



UNIVERSIDADE DO VALE DO TAQUARI – UNIVATES

CURSO DE ENGENHARIA DE SOFTWARE

**SEGMENTAÇÃO DO DISCO ÓPTICO E ESCAVAÇÃO COM O USO
DE REDES NEURAIS PARA AUXÍLIO À DETECÇÃO DE
GLAUCOMA**

Eduardo Caron

Lajeado, junho de 2020

Eduardo Caron

**SEGMENTAÇÃO DO DISCO ÓPTICO E ESCAVAÇÃO COM O USO
DE REDES NEURAIIS PARA AUXÍLIO À DETECÇÃO DE
GLAUCOMA**

Monografia apresentada ao Centro de Ciências Exatas e Tecnológicas da Universidade do Vale do Taquari - Univates, como parte dos requisitos para a obtenção do título de Bacharel em Engenharia de Software.

Orientador: Prof. Dr. Evandro Franzen

Lajeado, junho de 2020

AGRADECIMENTOS

A minha família, meu pai Edemilson A. Caron, minha mãe Marilene B. Caron e meu irmão Henrique Caron por serem essenciais na minha vida e por sempre me apoiarem.

A minha namorada Gabrielli Sartori pela compreensão, apoio e por não medir esforços para me ajudar durante o período e todos os fins de semana dedicados aos estudos.

Ao meu professor orientador Evandro Franzen pela disponibilidade, instruções e dedicação durante o período de orientação.

Aos meus colegas de TCC por compartilharem momentos de angústia e apoio.

Aos meus grandes amigos, que permitiram que essa caminhada fosse mais alegre.

E a todos os professores, por todo o conhecimento e ajuda durante os meus estudos e elaboração do meu TCC.

RESUMO

O uso de Inteligência Artificial (IA) para resolução de problemas complexos vem ganhando força no mundo inteiro. Técnicas baseadas em IA estão sendo amplamente adotadas no reconhecimento de imagens, de fala e de processamento de linguagem natural em diferentes áreas. A área da saúde está aderindo cada vez mais a Inteligência Artificial para diagnóstico e classificação de doenças, tendo como exemplo a doença de glaucoma, que pode ser detectada tanto por meio do processamento automatizado de grandes conjuntos de dados, quanto pela detecção precoce de novos padrões de doenças. Além disso, a IA promete mudar fundamentalmente as pesquisas que visam entender o desenvolvimento, a progressão e o tratamento do glaucoma, identificando novos fatores de risco e avaliando a importância dos existentes. Diante disso, o objetivo do presente trabalho foi verificar se a Rede Neural Artificial (RNA) de arquitetura U-Net é capaz de realizar a segmentação do disco óptico e a segmentação da escavação a partir de imagens de fundo de olho da base de imagens RIM-ONE, servindo de apoio aos profissionais na detecção do glaucoma. Como resultado, foi obtido um *F1-score* médio de 0,9440 para a segmentação do disco óptico e um *F1-score* médio de 0,8350 para a segmentação da escavação, comprovando que a rede neural U-Net é capaz de realizar a segmentação com boas métricas.

Palavras-chave: Glaucoma. Inteligência Artificial. Rede Neural Artificial. Segmentação de imagem. U-Net.

ABSTRACT

The use of Artificial Intelligence (AI) to solve complex tasks has been gaining force all over the world. AI-based techniques are being widely adopted in image recognition, speech recognition and natural language processing in different areas. The health area is increasingly adhering to Artificial Intelligence for diagnosis and classification of diseases, with the example of the glaucoma disease, which can be detected either through the automated processing of large datasets, or by the early detection of new patterns of diseases. In addition, AI promises to fundamentally change the research aimed at understanding the development, progression and treatment of glaucoma, identifying new risk factors and assessing the importance of existing ones. Therefore, the objective of the present work was to verify if the Artificial Neural Network (ANN) using the U-Net architecture is able to perform optic disk segmentation and optic cup segmentation on fundus images of the RIM-ONE image dataset, supporting professionals in the detection of glaucoma. As a result, an average F1-score of 0.9440 was obtained for the optic disk segmentation and an average F1-score of 0.8350 for the optic cup segmentation, proving that the U-Net neural network is capable of performing segmentation with good metrics.

Keywords: Artificial Intelligence. Artificial Neural Network. Glaucoma. Image segmentation. U-Net.

LISTA DE FIGURAS

Figura 01 – Estrutura do olho	19
Figura 02 – Imagem de fundo do olho.....	21
Figura 03 – Disco óptico e escavação na direção vertical.....	22
Figura 04 – Progressão do glaucoma	23
Figura 05 – Componentes do neurônio biológico	31
Figura 06 – Elementos básicos de um neurônio	32
Figura 07 – Exemplos de arquiteturas de RNAs	34
Figura 08 – Rede MLP típica com duas camadas intermediárias	36
Figura 09 – Fluxos de sinal de uma rede perceptron de múltiplas camadas.....	37
Figura 10 – Processo de convolução de uma RNC.....	40
Figura 11 – Aplicação de <i>max pooling</i> em imagem 4x4 utilizando um filtro 2x2... 41	
Figura 12 – Imagem original RIM-ONE R3 e suas máscaras.....	50
Figura 13 – Arquitetura de uma rede U-Net	54
Figura 14 – Fluxo das atividades.....	55
Figura 15 – Imagem de fundo de olho original e após pré-processamento.....	56
Figura 16 – Canais em uma imagem RGB do fundo do olho	57
Figura 17 – Exemplos do resultado de aumento dos dados.....	59
Figura 18 – Imagem do fundo do olho e máscara do disco óptico	63
Figura 19 – Comparação após corte da área de interesse com corte original	64
Figura 20 – Imagem do fundo do olho cortada na área de interesse e máscara do disco óptico	67

Figura 21 – Resultados e comparação da segmentação do disco óptico na imagem Glaucoma 5	69
Figura 22 – Resultados e comparação da segmentação do disco óptico na imagem Glaucoma 2	70
Figura 23 – Imagem do fundo do olho cortada na área de interesse e máscara da escavação	72
Figura 24 – Resultados e comparação da segmentação da escavação na imagem Glaucoma 5	74
Figura 25 – Resultados e comparação da segmentação da escavação na imagem Glaucoma 4	75

LISTA DE TABELAS

Tabela 01 – Comparação de resultados da segmentação do disco óptico entre autores	71
Tabela 02 – Comparação de resultados da segmentação da escavação entre autores	76

LISTA DE QUADROS

Quadro 01 – Matriz confusão	26
Quadro 02 – Resumo dos trabalhos relacionados e resultados obtidos	46
Quadro 03 – Resultados da segmentação de disco óptico na base de dados RIM-ONE R3.....	68
Quadro 04 – Resultados da segmentação da escavação na base de dados RIM-ONE R3.....	73

LISTA DE ABREVIATURAS E SIGLAS

AG	Algoritmo genético
AI	<i>Artificial Intelligence</i>
ANN	<i>Artificial Neural Network</i>
API	<i>Application Programing Interface</i>
CDR	<i>Cup Disk Ratio</i>
CLAHE	<i>Contrast-Limited Adaptive Histogram Equalization</i>
CNN	<i>Convolutional Neural Networks</i>
CPU	<i>Central Processing Unit</i>
FN	<i>False negative</i>
FP	<i>False positive</i>
GPU	<i>Graphics Processing Unit</i>
IA	Inteligência Artificial
MCP	<i>McCulloch Pitts</i>
MLP	<i>Multi-Layer Perceptron</i>
OD	<i>Optical Disc</i>
OMS	Organização Mundial de Saúde
ONEIROS	<i>Open-ended Neuro-Electronic Intelligent Robot Operating System</i>
ONH	<i>Optic Nerve Head</i>
PIL	<i>Python Imaging Library</i>
PPV	<i>Positive Predictive Value</i>
ReLU	<i>Rectified Linear Units</i>
RF	Requisito Funcional

RGB	<i>Red Green Blue</i>
RIM-ONE	<i>Retinal Image database for Optic Nerve Evaluation</i>
RNA	Rede Neural Artificial
RNC	Redes Neurais Convolucionais
RNF	Requisito Não Funcional
TP	<i>True Positive</i>
TN	<i>True Negative</i>
ULR	Unidade Linear Retificada
VLSI	<i>Very Large Scale Integration</i>

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Objetivos	16
1.1.1 Objetivo geral.....	Erro! Indicador não definido.
1.1.2 Objetivos específicos.....	Erro! Indicador não definido.
1.2 Estrutura do trabalho	17
2 REFERENCIAL TEÓRICO.....	18
2.1 O olho humano	18
2.2 O glaucoma	20
2.3 Inteligência Artificial	23
2.3.1 Aprendizado de máquina	24
2.3.1.1 Avaliação do desempenho de algoritmos de classificação	26
2.3.1.1.1 Precisão	28
2.3.1.1.2 Revocação.....	28
2.3.1.1.3 Medida F	28
2.4 Redes Neurais Artificiais	28
2.4.1 Neurônio Biológico	30
2.4.2 Neurônio artificial.....	32
2.4.3 Arquiteturas de rede	33
2.4.4 Perceptron	35
2.4.4.1 Perceptron com múltiplas camadas	35
2.4.4.2 <i>Backpropagation</i>	38

2.4.5 Redes Neurais Convolucionais	38
2.4.5.1 Convolução	39
 3 TRABALHOS RELACIONADOS	 43
 4 MATERIAIS E MÉTODOS	 47
4.1 Classificação da pesquisa	47
4.2 Tecnologias	48
4.2.1 Equipamentos	48
4.2.2 Bancos de imagens	49
4.2.2.1 RIM-ONE R3	49
4.2.3 Python	50
4.2.3.1 Bibliotecas python	50
4.2.3.1.1 TensorFlow	51
4.2.3.1.2 Keras	52
4.2.4 Arquitetura de rede neural U-Net	53
4.3 Desenvolvimento	54
4.3.1 Pré-processamento	55
4.3.1.1 Utilização do canal G	57
4.3.1.2 Aumento dos dados	58
4.3.2 Arquitetura da rede neural	59
4.3.3 Configuração do treinamento	60
4.3.4 Busca automatizada da área de interesse	62
 5 RESULTADOS E DISCUSSÕES	 66
5.1 Segmentação do disco óptico	66
5.2 Segmentação da escavação	71
 6 CONCLUSÕES	 77
6.1 Sugestões de desenvolvimentos futuros	78
 REFERÊNCIAS	 79

ANEXOS	85
ANEXO A – Resultados da segmentação do disco óptico	86
ANEXO B – Resultados da segmentação da escavação.....	87

1 INTRODUÇÃO

Os dados revelados pela Organização Mundial de Saúde (OMS) em 2010 sobre deficiências de visão no mundo revelam que o número estimado de pessoas com deficiências visuais é de 285 milhões, 39 milhões de pessoas são cegas e 246 milhões possuem baixa visão. As doenças que são maiores causadoras de perda de visão são catarata, 51%; glaucoma, 8%; degeneração macular relacionada à idade, 5%; cegueira na infância e opacidade corneana, 4% (PASCOLINI e MARIOTTI, 2011).

A degradação das fibras nervosas danifica gradualmente o nervo óptico, causando a doença crônica de glaucoma (MARY et al., 2016). Segundo Giangiacomo e Coleman (2009) glaucoma é ranqueado como a segunda maior causa de cegueira no mundo, assim como a maior causa de perda de visão. Além disso, o número de indivíduos afetados por glaucoma tem projeção de atingir 80 milhões de pessoas até o ano de 2020.

O estágio inicial do glaucoma é assintomático e, segundo as pesquisas de Giangiacomo e Coleman (2009), Mitchell et al. (1996) e Michelson et al. (2008), 50% dos pacientes não estão cientes de que estão com a doença. A progressão do glaucoma sem diagnóstico e tratamento adequado, gradualmente leva o paciente a perda irreversível da visão (ALMAZROA et al., 2015). Conforme Michelson et al. (2008), o tratamento do glaucoma em estágios iniciais pode reduzir as chances de perda total e irreversível da visão, após 20 anos do início das medicações em 50%.

Assim sendo, o diagnóstico precoce do glaucoma é essencial para o gerenciamento ideal da doença.

A detecção precoce do glaucoma com a ajuda de imagens de fundo de olho é uma importante etapa no tratamento, pois permite que os especialistas possam tomar decisões contra a doença, a fim de garantir ao paciente uma melhor qualidade de vida (CHEN et al., 2015).

A coleta de imagens de fundo do olho é uma das principais técnicas utilizadas no diagnóstico do glaucoma por ser não-invasiva (NAYAK et al., 2009). O profissional de saúde utiliza uma câmera capaz de extrair características importantes das imagens de fundo do olho, como os vasos sanguíneos e o disco ótico (OD, do inglês *optic disc*). Estas estruturas são capazes de fornecer informações úteis para o diagnóstico do glaucoma (NAYAK et al., 2009).

Na década de 60, Armaly (1967) observou que o tamanho da escavação relativa ao tamanho do disco óptico, ou a proporção escavação disco (*cup disk ratio*, CDR), era uma medida útil para cálculo de danos causados pelo glaucoma. As imagens do fundo do olho são utilizadas pelos profissionais da saúde para demarcação e cálculo da proporção de CDR. Esta demarcação é feita manualmente, tornando-se uma tarefa desafiadora e trabalhosa devido a quantidade de vasos e alterações graduais de cores no olho.

Visto que a segmentação manual é uma tarefa demorada e altamente subjetiva para humanos, muitas técnicas e algoritmos para automatização desta tarefa vem sendo desenvolvidos, vários deles com o uso de aprendizado profundo e redes neurais. Segundo Shallue (2018), o uso de redes neurais em anos recentes permitiu o desenvolvimento de um grande número de soluções para reconhecimento de padrões, como visão de computador, reconhecimento de imagem e reconhecimento de fala.

Uma Rede Neural Convolucional (CNN, do inglês *Convolutional Neural Networks*), é uma técnica de aprendizagem profunda que vem tendo grande visibilidade nos últimos anos. Segundo Sevastopolsky (2017), além de ser uma técnica que reduz os esforços de trabalho, por ser automatizada, a grande

vantagem do uso das CNNs está na universalidade, já que a mesma rede geralmente pode reconhecer vários padrões em diferentes imagens e objetos através das características extraídas automaticamente, pois são capazes de inferir uma representação hierárquica das imagens para discriminação entre os padrões estabelecidos.

Considerando a aplicabilidade das redes neurais, o trabalho descrito neste documento propôs a aplicação de redes neurais convolucionais profundas para a realização da segmentação do disco óptico e da escavação através da análise de imagens do fundo do olho. Para estas tarefas, foram utilizadas imagens de domínio público de bases de dados RIM-ONE. Estas imagens foram a base para treinamento da rede neural convolucional com o auxílio do TensorFlow, utilizando uma rede neural de arquitetura U-Net. Para validação, os resultados foram comparados com a classificação manual feita por médicos e também comparados com métodos já existentes.

1.1 Objetivos

O objetivo geral do presente estudo é investigar se a rede neural de arquitetura U-Net é capaz de realizar a segmentação do disco óptico e escavação a partir de imagens de fundo do olho, a fim de apoiar a detecção da doença de glaucoma.

Os objetivos específicos do presente trabalho são:

- Desenvolver um procedimento de apoio à preparação de imagens para serem utilizadas em rede neural;
- Utilizar tecnologias e técnicas que facilitem o desenvolvimento e a experimentação da rede neural;
- Investigar se por meio de análise de imagens de fundo de olho, a rede neural de arquitetura U-Net é capaz de realizar a segmentação do disco óptico e da escavação.

1.2 Estrutura do trabalho

Para auxiliar no entendimento, o presente trabalho está estruturado em seis capítulos e foi organizado da maneira descrita no parágrafo a seguir.

O capítulo dois contém o referencial teórico, que foi utilizado como base para o desenvolvimento do trabalho, conta com detalhes técnicos sobre temas como o olho humano, inteligência artificial e redes neurais artificiais. O capítulo três descreve ferramentas, métodos e resultados utilizados em cinco pesquisas diferentes, onde foram realizadas técnicas relacionadas ao presente documento. No capítulo quatro é apresentada a metodologia que foi utilizada para realizar o presente trabalho e o método científico que ele se enquadra. Além disso, apresenta as bases de dados, ferramentas e procedimentos técnicos utilizados para auxílio nesta pesquisa. O capítulo cinco apresenta os resultados obtidos pela aplicação dos métodos. Já o capítulo seis, apresenta as conclusões do trabalho.

2 REFERENCIAL TEÓRICO

O referencial teórico tem como objetivo explicar conceitos utilizados neste trabalho, trazendo fundamentos de autores para explicar as teorias gerais sobre cada assunto. Os assuntos abordados iniciam com o olho humano e passam pela doença de glaucoma, para assim introduzir o leitor ao problema. Posteriormente é mostrado sobre IA e Redes Neurais, assuntos importantes discutidos neste trabalho.

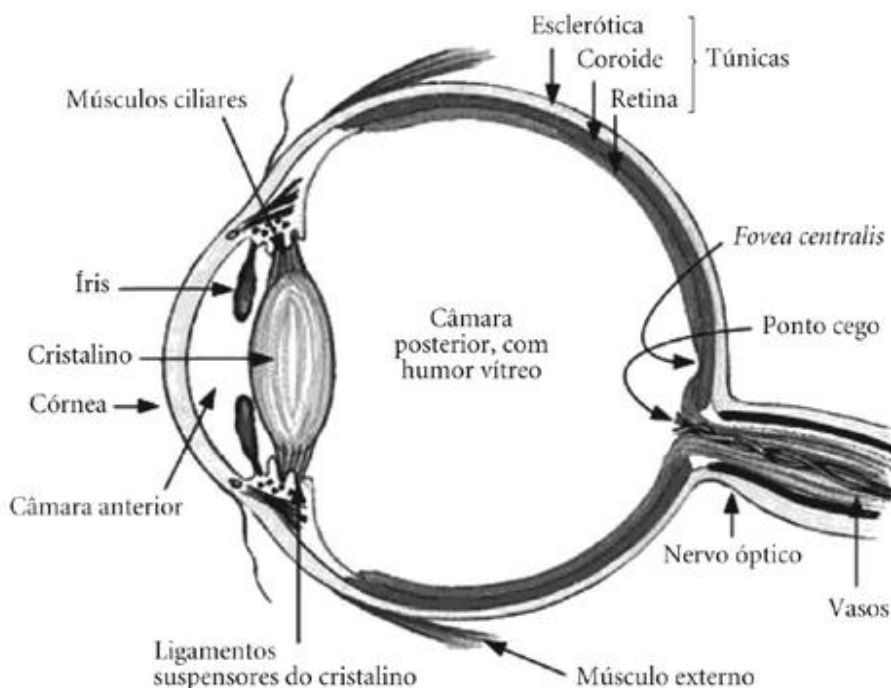
2.1 O olho humano

Conhecer a anatomia do olho humano é importante para este estudo devido a necessidade do entendimento das partes internas e assim, facilitando o desenvolvimento das técnicas que este estudo tem por objetivo.

O olho humano é um complexo sistema sensorial que, em conexão com o cérebro, desempenha funções de formação e processamento de imagens. Ele é responsável por captar os raios luminosos refletidos pelos objetos, que após atravessar o globo ocular formam uma imagem invertida na retina, e então, transportar esta imagem no formato de impulsos elétricos até o cérebro, onde este interpreta a imagem na posição correta. (PASSOS, ANDRADE e LEMAIRE, 2008).

A Figura 01 faz a representação dos órgãos responsáveis pela formação das imagens, que estão descritos conforme Courrol e Preto (2011) nos parágrafos a seguir.

Figura 01 – Estrutura do olho



Fonte: Courrol e Preto (2011).

A córnea, que faz parte da camada externa do olho, tem como função transmitir, realizar refração da luz e proteger a camada anterior do globo ocular. Funciona como uma lente e, junto ao cristalino, compõem o sistema responsável por focalizar imagens na retina.

A membrana delgada e transparente que cobre a parte branca do olho e a parte interna das pálpebras é chamada de conjuntiva. Sua função é defender a superfície ocular contra influências externas e realizar lubrificação ocular.

A íris, que dá a cor aos olhos, controla a intensidade de luz que entra no olho. Os raios passam para o interior do globo ocular através de uma pequena abertura no seu centro, chamada de pupila. A pupila varia de tamanho conforme a intensidade de luz, dilatando-se quando há pouca luz para que mais luz chegue até

a retina e contraindo-se em momentos onde existir mais luz, diminuindo assim a intensidade de luz que chega ao interior do olho.

A parte branca do olho é chamada de esclera ou esclerótica. Sua responsabilidade é a manutenção do formato do olho e a proteção de suas estruturas, em decorrência de ter uma consistência mais rígida. A função do cristalino é ajustar o foco nos objetos conforme a distância que eles se situam do olho, assim permitindo a visão de objetos mais próximos ou distantes.

Ainda conforme Courrol e Preto (2011), a retina é um fundamental tecido do olho onde a imagem é diretamente focada. Suas tarefas são receber, focar e enviar as imagens para o cérebro. Na mácula as imagens se formam com mais nitidez, e ficam mais nítidas na região central da mácula, chamada de fóvea. A continuação das células nervosas que compõem a retina é denominada nervo óptico. Sua responsabilidade é levar para o cérebro as imagens captadas na retina, onde a visão será formada.

2.2 O glaucoma

O glaucoma é um grupo de distúrbios nervosos ópticos que têm em comum a contínua degeneração das células glanglorianas da retina e seus axônios, resultando em uma desfiguração de uma parte do nervo óptico (ONH, do inglês, *optic nerve head*), conhecida como escavação e, conseqüentemente, a perda de visão (WEINREB e KHAW, 2004). Segundo Pascolini e Mariotti (2011), no relatório da OMS sobre deficiências de visão de 2010, o glaucoma é a terceira maior causa de cegueira no mundo.

A causa do glaucoma é relacionada ao humor aquoso, um fluído que circula na parte interna do olho, cuja função é nutrir a área da íris e córnea e exercer pressão para manter o formato do olho. Este fluído é produzido continuamente, causando aumento da pressão denominada pressão intraocular. Quando não ocorre uma drenagem suficiente devido a uma obstrução no canal de drenagem,

pode assim resultar em pressão alta na região, danificando e mudando a aparência do nervo óptico (NÄTTINEN, 2015).

O dano causado pelo glaucoma é normalmente associado a elevada pressão intra-ocular. O dano à ONH acarreta a perda parcial ou total do campo de visão. Danos gerados ao campo de visão são irreversíveis, entretanto, com o devido tratamento nos estágios iniciais do glaucoma, a perda pode ser adiada. Se não for devidamente tratada, o dano causado ao campo visual normalmente piora e se espalha, podendo chegar ao ponto da perda completa da visão (WEINREB e KHAW, 2004).

Retinografia, fotografia do nervo óptico ou (fotografia) imagem do fundo de olho é a mais antiga e estabelecida forma de capturar estruturas do olho como fóvea, vasos sanguíneos e disco óptico. A Figura 02 representa uma imagem de fundo de olho capturada por retinografia. Este procedimento é valorizado pois permite que seja realizada avaliação de várias características do glaucoma que ainda não são possíveis por métodos semi-automáticos. A avaliação da progressão do glaucoma em imagens de fundo de olho é subjetiva e depende da experiência do profissional clínico (KUMMET, 2013).

Figura 02 – Imagem de fundo do olho

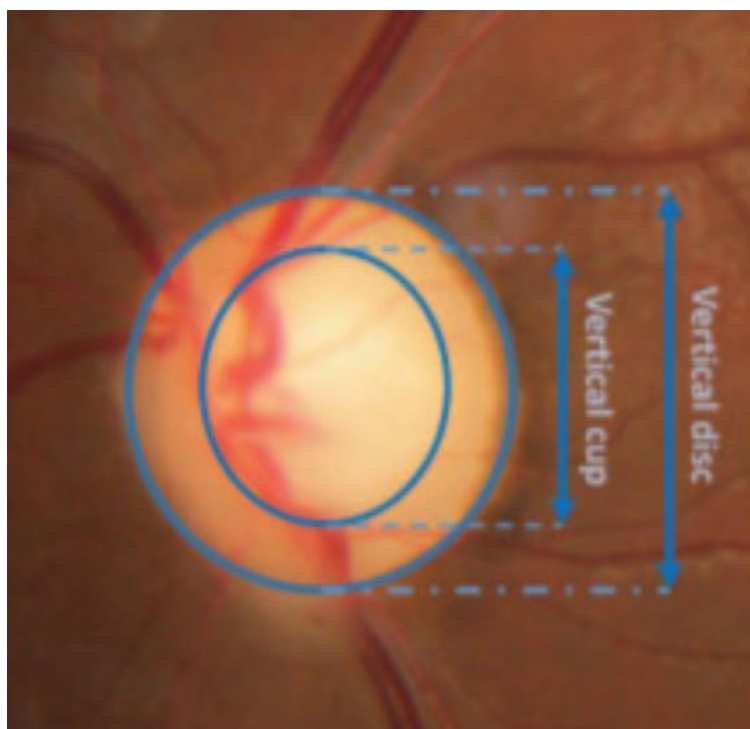


Fonte: Almazroa et. al (2018).

A proporção CDR, que mede o diâmetro vertical da escavação e do disco óptico está entre os exames de imagem mais comuns para detecção de glaucoma

(KUMAR et al., 2007). Geralmente, uma proporção CDR de 0,5 é considerada suspeita de glaucoma e quanto maior a proporção, maior a probabilidade de ser considerado glaucoma (GARWAY-HEATH et al., 1998). A Figura 03 mostra a medição do CDR a partir de uma imagem de fundo de olho.

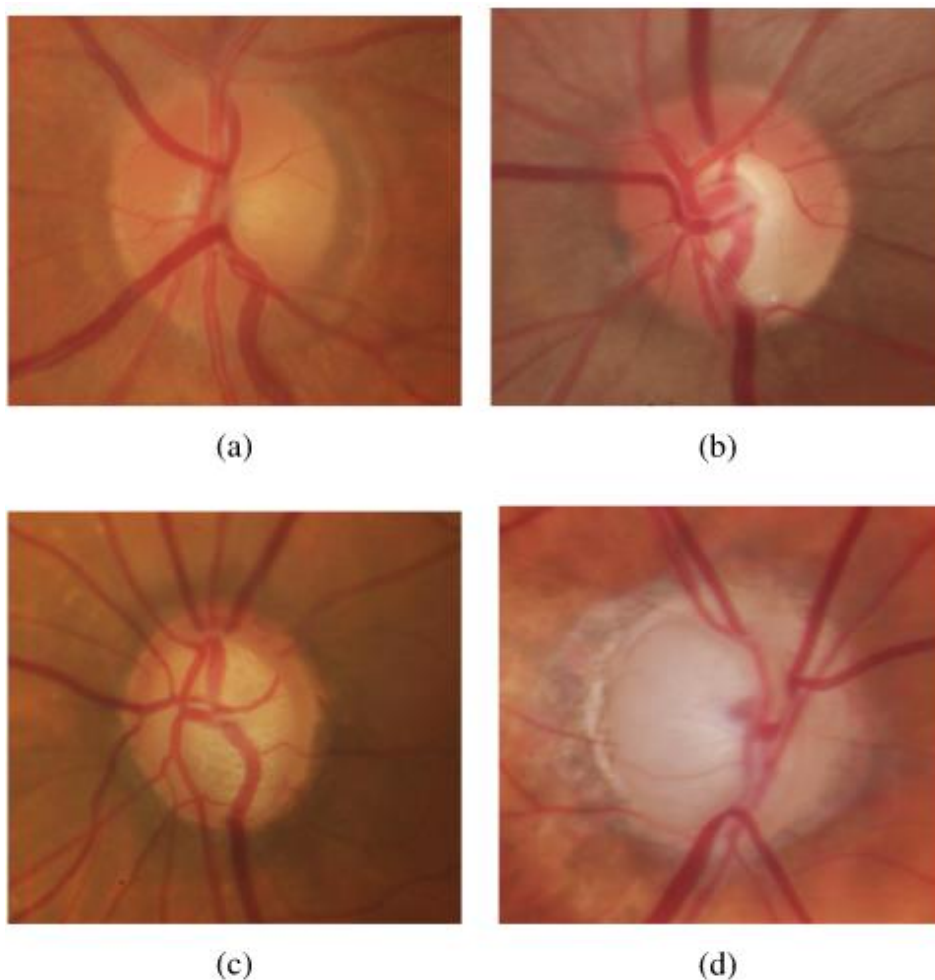
Figura 03 – Disco óptico e escavação na direção vertical



Fonte: Ruengkitpinyo et al. (2015).

A Figura 04 ilustra, em imagens de fundo do olho, diferentes estágios do glaucoma à medida que progride, onde em (a), olho sem glaucoma; (b), glaucoma precoce; (c) glaucoma moderado; (d) glaucoma profundo.

Figura 04 – Progressão do glaucoma



Fonte: Mary; Rajsingh e Naik (2016).

2.3 Inteligência Artificial

Coppin (2013), afirma que a inteligência artificial pode ser definida como o estudo de sistemas que, ao serem observados, causam a impressão de serem inteligentes. Esta definição é boa, mas não é o suficiente para descrever todos os campos que a inteligência artificial é capaz de agir. Métodos de IA são usados para a resolução de problemas desde muito simples até extremamente complexos. Com estas características, Coppin (2013) sugere assim que outra definição pode ser usada, afirmando então que IA envolve o uso de métodos inspirados no comportamento de seres humanos e outros animais para resolução de problemas complexos.

Métodos de IA são geralmente divididos entre as subáreas que resolvem problemas práticos comuns, porém, ao mesmo tempo difíceis e que requerem uma diferente gama de ferramentas e habilidades. Algumas áreas mais proeminentes são a robótica, reconhecimento de fala, visão computacional, processamento de linguagem natural e o aprendizado de máquina (KAPLAN, 2016).

O objetivo da IA não é mais criar robôs inteligentes como humanos, mas sim usar técnicas como algoritmos, heurísticas e metodologias que se baseiam nas maneiras que o cérebro humano resolve problemas, explica Coppin (2013). Assim, sistemas foram projetados como a Analogia de Thomas Evans e a Arquitetura *Copycat* de Melanie Mitchell, os quais foram projetados para serem capazes de resolver problemas que envolvem analogias.

2.3.1 Aprendizado de máquina

Aprendizado de máquina (*machine learning*) é um conjunto de técnicas que utilizam algoritmos para realizar análises em grandes bases de dados. O aprendizado de máquina é capaz de realizar análise preditiva muito mais rapidamente que qualquer humano, resultando assim em um trabalho humano mais eficiente (MUELLER, 2016).

Para Mueller (2016), a parte essencial para entendimento da definição do aprendizado de máquina é que este proporciona apenas a parte de aprendizado de uma IA. O aprendizado de máquina participa do contexto de IA pela realização de tarefas como adaptar-se às novas circunstâncias não previstas inicialmente pelo desenvolvedor, detectar padrões em dados, criar novos comportamentos baseados nos padrões detectados e tomar decisões baseados nos fracassos e sucessos do que foi aprendido com os dados.

O processo de aprendizagem pode ser classificado em dois tipos, aprendizagem supervisionada e aprendizagem não-supervisionada.

Algoritmos de aprendizado supervisionado utilizam uma base de dados que além de conter características, contém um rótulo ou classe. Uma base de dados de

flores do gênero íris, por exemplo, é anotada com as espécies de cada planta. Um método de aprendizado supervisionado, pode então, estudar essa base de dados e aprender a classificar, baseado em suas medidas, cada planta em sua espécie correta. Algoritmos de aprendizado não supervisionado utilizam uma base de dados que contém várias características e então aprendem propriedades da estrutura da base de dados (GOODFELLOW et al., 2016).

Técnicas não supervisionadas empregam a observação de uma grande quantidade de exemplos em um vetor aleatório x para tentar, implicitamente ou explicitamente, aprender a probabilidade da distribuição $p(x)$, ou algumas propriedades úteis desta distribuição. Já as técnicas supervisionadas empregam a observação de vários exemplos de um vetor aleatório x e de um valor associado y , e então aprender a prever y pelo x estimando $p(y|x)$. Na aprendizagem supervisionada é utilizado um professor ou instrutor que mostra para o sistema o que fazer através do uso deste valor associado y , por isso se dá o nome de aprendizagem supervisionada. Este instrutor ou professor não está presente na aprendizagem não supervisionada, o seu algoritmo deve aprender a gerar resultados satisfatórios para os dados sem ser guiado (GOODFELLOW et al., 2016).

Várias tarefas podem ser resolvidas com uso de aprendizado de máquina. Segundo Goodfellow et al., (2016), algumas das mais comuns são:

- Classificação: o programa recebe uma entrada e este deve informar a quais categorias esta entrada pertence;
- Regressão: o programa recebe uma entrada e este deve prever um valor numérico referente a esta entrada. Diferente da classificação, este algoritmo apresenta uma função como saída;
- Transcrição: o programa recebe uma entrada com dados em formato relativamente não estruturado e este deve retornar dados em formato textual;
- Tradução de máquina: o programa recebe textos em uma linguagem e deve retornar o mesmo texto em outro idioma;

- Detecção de anomalias: o algoritmo recebe um conjunto de entradas e identifica quais destas entradas estão com características atípicas;
- Resumo e exemplificação: o programa é desenvolvido para gerar novos exemplos que são similares aos de seus dados de treinamento;
- Redução de ruídos: o algoritmo recebe como entrada um dado corrompido e deve retornar como saída uma previsão para a parte corrompida ou então uma distribuição de probabilidade.

2.3.1.1 Avaliação do desempenho de algoritmos de classificação

Conforme Han, Kamber e Pei (2012), uma forma de avaliação do desempenho de um algoritmo de classificação é medir sua capacidade preditiva. Esta tarefa deve ser feita com a exposição do modelo de classificação à dados não utilizados durante a etapa de treinamento, permitindo assim a identificação de ruídos e melhor generalização de dados.

Segundo Tan, Steinbach e Kumar (2009), na classificação de problemas binários, predicação entre duas classes, a técnica de matriz confusão é utilizada, onde as colunas representam as classes de previsão e as linhas as classes atuais dos dados, conforme ilustra a Quadro 01.

Quadro 01 – Matriz confusão

		Classe prevista	
		0	1
Classe atual	0	Verdadeiro negativo	Falso positivo
	1	Falso negativo	Verdadeiro positivo

Fonte: Adaptado pelo autor de Tan; Steinbach e Kumar (2009, p.351).

Os termos que compõem uma matriz de confusão, como no Quadro 01, conforme os autores possuem as seguintes definições:

- Verdadeiro positivo (TP): quantidade de exemplos positivos que foram corretamente classificados;
- Falso negativo (FN): quantidade de exemplos negativos incorretamente classificados;
- Falso positivo (FP): quantidade de exemplos positivos incorretamente classificados;
- Verdadeiro negativo (TN): quantidade de exemplos negativos corretamente classificados.

Existem ainda outras métricas importantes para medição do desempenho de classificadores, conforme listado a seguir (HAN; KAMBER; PEI, 2012):

- Acuracidade: taxa dos objetos de testes corretamente classificados;
- Taxa de erro: percentual de objetos incorretamente classificados;
- Especificidade: taxa de objetos falsos verdadeiros incorretamente classificados;
- Revocação (*recall*): taxa de objetos verdadeiro positivos corretamente classificados;
- Precisão (*precision*): taxa de objetos positivos corretamente classificados;
- Medida F (*F1-score*): média harmônica entre as métricas de precisão e revocação.

As seções seguintes introduzem de maneira detalhada importantes métricas de avaliações que foram utilizadas neste trabalho.

2.3.1.1.1 Precisão

A precisão (*precision*) pode ser pensada como uma medição de exatidão, ou seja, a porcentagem de valores positivos classificados como positivos (HAN; KAMBER; PEI, 2012). A Equação 01 representa como é medida a precisão.

$$precisão = \frac{\text{verdadeiro positivo (TP)}}{\text{verdadeiro positivo (TP)} + \text{falso positivo (FP)}} \quad 01$$

2.3.1.1.2 Revocação

Revocação (*recall*) é a medida de completude, ou seja, qual a porcentagem de valores positivos foi corretamente previstos (HAN; KAMBER; PEI, 2012). A Equação 02 representa como é medida a revocação.

$$revocação = \frac{\text{verdadeiro positivo (TP)}}{\text{verdadeiro positivo (TP)} + \text{falso negativo (FN)}} \quad 02$$

2.3.1.1.3 Medida F

A medida F (*F1-score*) é média harmônica da precisão e da revocação (HAN; KAMBER; PEI, 2012). Ela é também conhecida como coeficiente Dice (*dice coefficient*).

A Equação 03 representa como é feita a medida F.

$$F1\ score = 2 \times \frac{\text{precisão} \times \text{revocação}}{\text{precisão} + \text{revocação}} \quad 03$$

2.4 Redes Neurais Artificiais

Redes neurais artificiais (RNA) são redes computacionais que tentam simular, em uma maneira geral, o processo de decisão das redes de células nervosas (neurônios) do sistema nervoso central biológico. Esta simulação se

baseia no conhecimento neurofisiológico dos neurônios biológicos e das redes desses neurônios (GRAUPE, 2013).

Haykin (2008) explica que uma rede neural é como uma máquina desenvolvida para modelar a maneira que o cérebro realiza uma tarefa ou função específica. A rede é implementada com o uso de componentes eletrônicos ou é simulada digitalmente com o uso de programação em um computador. Para atingir um bom desempenho, as redes neurais utilizam interligações maciças de células computacionais conhecidas como neurônios ou unidades de processamento.

RNAs são sistemas compostos de um agrupamento de unidades de processamento simples (nodos) contando assim com uma capacidade de paralelização e com característica de distribuição, afirmam Braga, Carvalho e Ludermir (2011). Estas unidades são interligadas por um grande número de conexões e compostas de uma ou mais camadas. Geralmente estas estruturas de conexões estão associadas a pesos que tem a função de armazenar o conhecimento representado no modelo e realizar a ponderação das entradas de cada neurônio.

As redes neurais artificiais permitem o uso de operações computacionais simples (adições, multiplicações e outros elementos de lógica fundamental) para resolver problemas matemáticos complexos. Uma rede neural artificial será computacionalmente e algoritmicamente simples e terá funcionalidades de auto-organização para permitir que ela seja capaz de resolver uma grande e distinta quantidade de problemas (GRAUPE, 2013).

Como afirma Haykin (2008), uma RNA extrai seu poder computacional através de sua estrutura paralelamente distribuída e de sua habilidade de aprender e, então, de generalizar. A generalização ocorre devido a RNA produzir saídas adequadas as entradas não informadas durante o seu treinamento (aprendizagem). Estas duas propriedades para o tratamento de dados fazem com que as RNAs tenham capacidade de resolver problemas complexos e de grande escala que até hoje são considerados intratáveis.

Haykin (2008) explica que o algoritmo de aprendizagem de uma RNA envolve operações de conexão entre os neurônios para a aquisição de conhecimento, chamadas de pesos sinápticos, cuja função é alterar os pesos sinápticos da rede para almejar o objetivo esperado.

A utilização de uma RNA proporciona propriedades e características referente ao conhecimento, como: Permite resolução de problemas lineares e não lineares; Aprende construindo um mapeamento de entrada-saída para um problema específico; Pode se adaptar para mudanças e aprender gradativamente junto com a mudanças dos dados; Além de aprender categorizar, pode fornecer uma taxa de confiança para a categoria selecionada; Sua estrutura interligada permite que ocorra um tratamento natural das informações contextuais; Suas características modularizadas em neurônios possibilita tolerância a falhas e processamento paralelo em grande escala; As suas teorias e algoritmos de aprendizagem podem ser empregados para diferentes objetivos com o uso de redes neurais; Uma interligação homogênea de módulos permite que redes modulares possam ser construídas (HAYKIN, 2008).

O desenvolvimento das RNAs é motivado pela inspiração com o cérebro, que é um exemplo vivo do paralelismo e de tolerância a falhas. Esta prova viva demonstra que é fisicamente possível ser desenvolvida uma ferramenta com tais características, tornando-a uma ferramenta altamente poderosa. Os neurobiólogos analisam as RNAs como instrumento de pesquisa para entendimento de fenômenos neurobiológicos. Por outro lado, os engenheiros utilizam a neurobiologia como inspiração para técnicas de resolução de problemas complexos (HAYKIN, 2008).

2.4.1 Neurônio Biológico

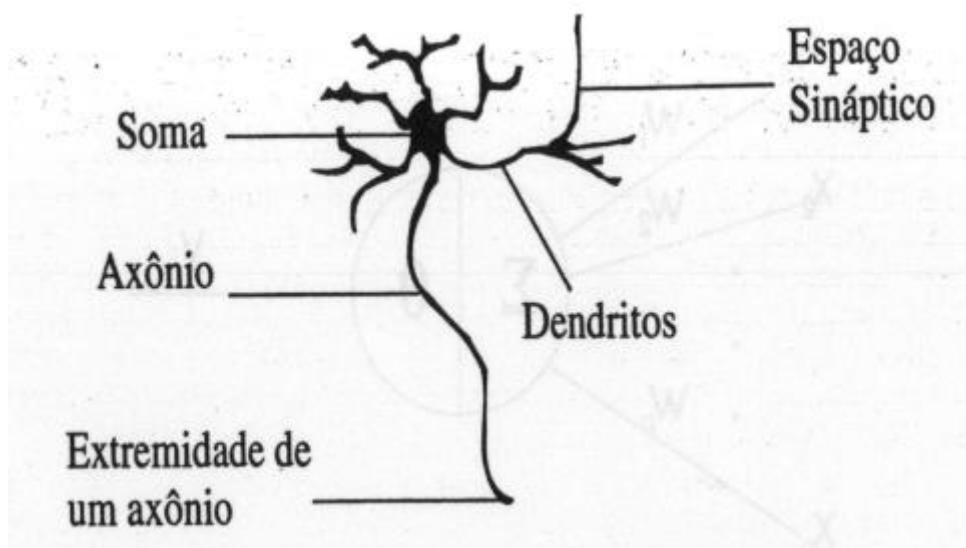
O cérebro humano é constituído por 10^{11} neurônios, sua célula mais importante. Cada neurônio mantém-se em processamento contínuo e se comunica com diversos outros neurônios. A sua estrutura, maneira como estão interligados e

seu funcionamento em conjunto, foram a base para estudo de RNAs, conforme Braga, Carvalho e Ludermir (2011).

Nem todo o funcionamento do cérebro foi desvendado pelo homem, que segue em estudo contínuo e atualização dos modelos de funcionamento. O que se tem entendimento até o momento é sua estrutura fisiológica básica, sendo está a base para o funcionamento das RNAs.

Os neurônios são simples elementos de processamento. Cada neurônio é composto pela soma, que é o corpo da célula, os dendritos e o axônio, como na Figura 05 (COPPIN, 2013).

Figura 05 – Componentes do neurônio biológico



Fonte: Braga; Carvalho e Ludermir (2011).

A função dos dendritos é receber as informações em forma de impulsos nervosos oriundas de outros neurônios e conduzi-las para o corpo celular, onde esta informação é processada e são gerados novos impulsos. Através do axônio estes impulsos são transmitidos para outros dendritos dos neurônios seguintes. A sinapse é a interligação entre a terminação axônica de um neurônio e o dendrito de outro, é nela onde ocorre a união funcional dos nodos, formando assim redes neurais, demonstram Braga, Carvalho e Ludermir (2011).

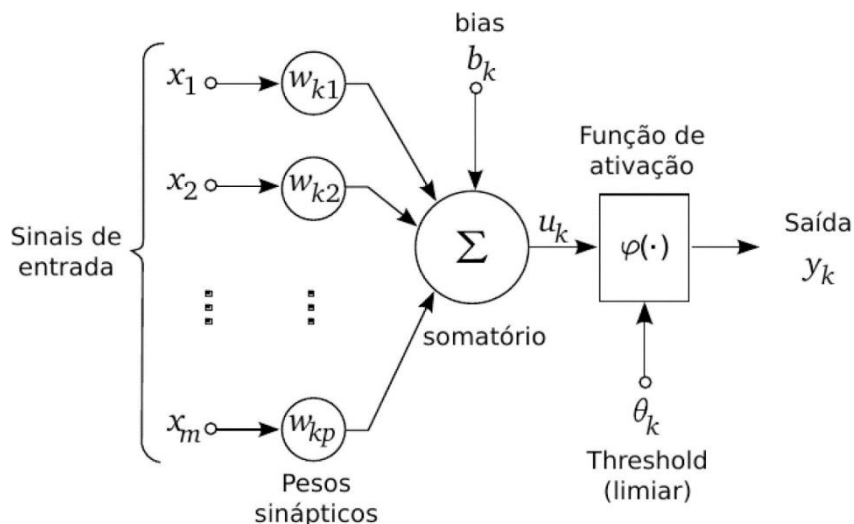
Copin (2013), afirma que mesmo cada neurônio sendo extremamente simples, a enorme e complexa rede de neurônios é capaz de processar um grande

e complicado volume de informação. Em termos de complexidade, o cérebro humano excede em muito qualquer dispositivo criado pelo homem ou, até mesmo, qualquer objeto ou estrutura ocorrendo naturalmente no universo até onde sabemos hoje.

2.4.2 Neurônio artificial

Um neurônio é uma unidade de processamento de informação fundamental para o funcionamento de uma RNA. A Figura 06 ilustra o modelo de um neurônio e identifica seus elementos básicos.

Figura 06 – Elementos básicos de um neurônio



Fonte: Adaptado pelo autor de Haykin (2008).

Os elementos de um neurônio artificial são assim caracterizados:

- Conjunto de sinapses, cada uma com peso ou força própria. Um sinal x_j na entrada da sinapse j conectada ao neurônio k é multiplicado pelo peso sináptico w_{kj} . A função do peso é determinar a importância desse sinal para o neurônio;
- Somador, que tem a função de agregar os sinais de entrada, sendo ponderado pelas sinapses do neurônio;
- Uma função de ativação que delimita a área de saída de um neurônio, geralmente $[-1, 1]$ ou $[0, 1]$.

Como afirma Coppin (2013), cada neurônio artificial de uma rede é chamado também de nó e possui um conjunto de entradas, além de uma função de ativação que é aplicada aos valores de entrada. Uma função de ativação é uma equação matemática que tem responsabilidade de restringir a saída do somador para que esta fique entre uma faixa de valores normalizada. A função limiar linear, ou função degrau é uma das funções de ativação mais populares dentre as existentes.

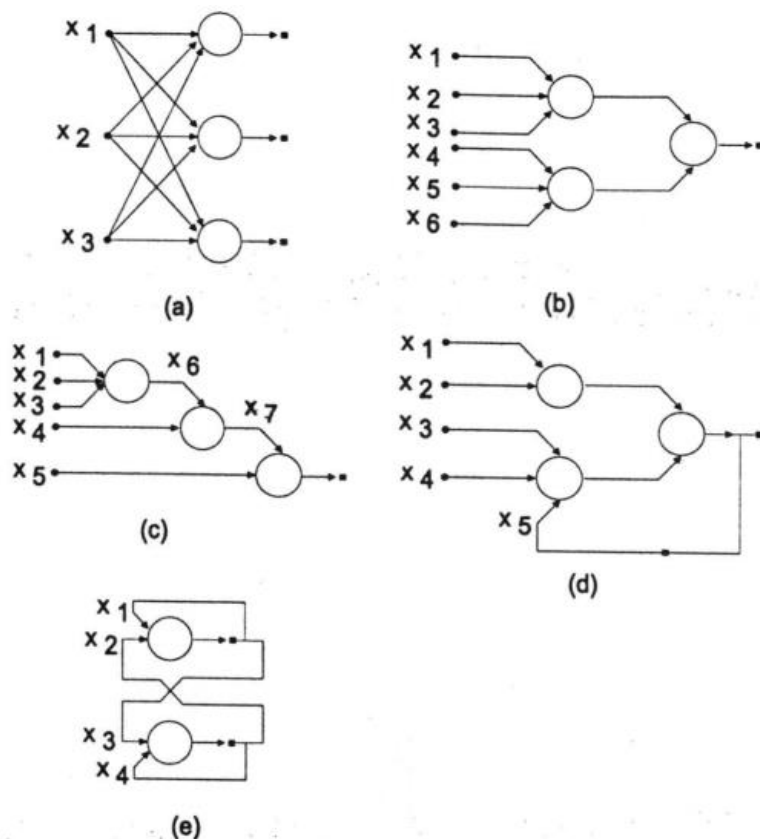
A função de ativação sigmóide (*sigmoid*) foi uma das funções utilizadas neste trabalho. A função de ativação sigmóide converte entradas de um intervalo de $-\infty$ até $+\infty$ para um nível de ativação do intervalo 0 até +1, explica Coppin (2013).

2.4.3 Arquiteturas de rede

Redes neurais com grandes quantidades de neurônios são usadas com frequência. A maneira de como os neurônios estão interligados determina como será construída e como será o funcionamento da rede. Sendo assim, definir qual arquitetura será usada é uma importante etapa inicial de projeto para o desenvolvedor da rede neural (MEHROTRA; MOHAN; RANKA, 1997).

Para Braga, Carvalho e Ludermir (2011), definir a arquitetura é importante para sua concepção pois ela restringe o tipo de problema que pode ser tratado pela RNA. Redes com uma única camada de nodos *McCulloch Pitts* (MCP), por exemplo, possuem capacidade apenas para resolver problemas separáveis linearmente. Já as redes recorrentes, são mais indicadas para resolver problemas na área de processamento temporal. Na definição da arquitetura, os parâmetros como número de camadas da rede, número de nodos por camada, topologia de rede e tipo de conexão entre os nodos devem ser considerados. A Figura 07 representa alguns tipos de arquiteturas de RNAs.

Figura 07 – Exemplos de arquiteturas de RNAs



Fonte: Braga; Carvalho e Ludermir (2011).

Relativo ao número de camadas, pode-se ter redes com camada única e redes com múltiplas camadas. Em redes com camada única, existe apenas um nó em cada entrada e cada saída, Figura acima 07 (a) e (e). Em redes com múltiplas camadas, existe mais de um nó entre alguma entrada e alguma saída, Figura 07 (b), (c) e (d).

Os tipos de conexões dos nodos podem ser *Feedback* ou *Feedforward*. No tipo de conexão *Feedback*, ou cíclica, a saída de um neurônio é usada como entrada em nodos de camadas de índice menor ou igual ao atual, como na Figura 07 (d) e (e). Já no tipo de conexão *feedforward*, ou acíclica, a saída de um neurônio não pode ser usada como entrada em nodos de camadas de índice menor ou igual ao atual, como na Figura 07 (a), (b) e (c).

2.4.4 Perceptron

As redes perceptron, que foram propostas por Rosenblatt em 1958, são redes formadas por um único neurônio com a capacidade de classificar suas entradas em duas categorias, permitindo que seja usado qualquer número de entrada. A entrada pode ser organizada em uma grade, podendo assim representar uma imagem ou um campo de visão com isso, permitindo que redes perceptron possam realizar tarefas de classificação de imagem e tarefas de reconhecimento (COPPIN, 2013).

De acordo com Haykin (2008), o perceptron é uma rede neural de formato mais simples usada na classificação de padrões lineares. Ele é composto basicamente por um neurônio único com pesos sinápticos ajustáveis e bias. O primeiro uso conhecido do algoritmo para ajustar os parâmetros de uma rede neural perceptron foi entre 1958 e 1962, em um procedimento de aprendizado feito por Rosemblatt, para uso em seu modelo cerebral. A tese proposta por Rosemblatt comprovou que se os vetores utilizados para treinamento do perceptron forem obtidos de duas classes linearmente separáveis, ocorre uma convergência no algoritmo e este então dispõe a área decisória em um hiperplano de duas classes. O perceptron construído com o uso de apenas um neurônio tem a limitação de apenas classificar padrões que possuem apenas duas hipóteses.

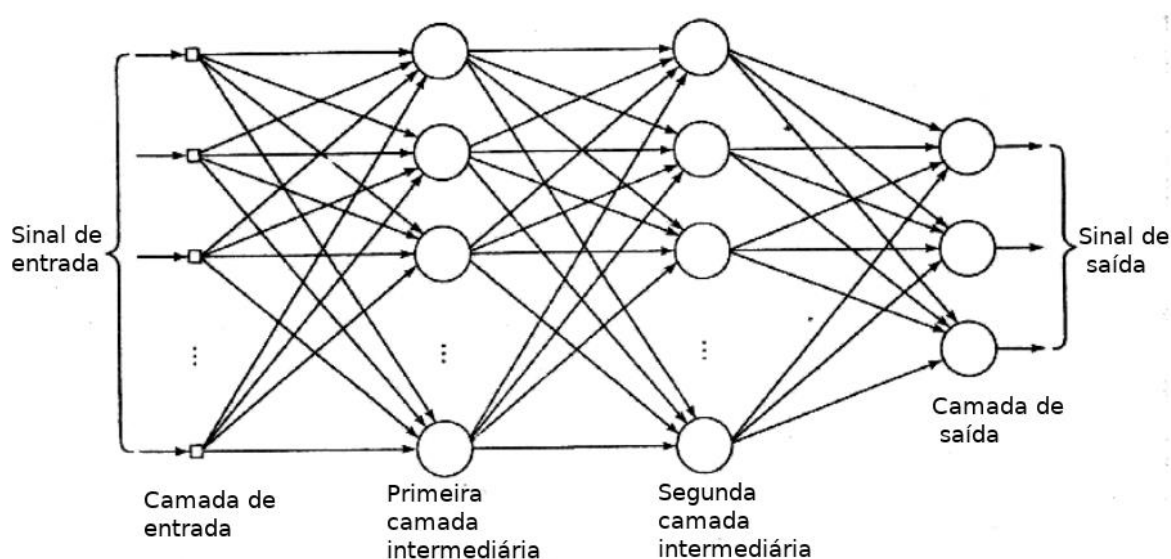
2.4.4.1 Perceptron com múltiplas camadas

As redes com múltiplas camadas de perceptrons, (*multi-layer perceptron*, MLP) representam uma generalização do perceptron de uma única camada. Estas redes são normalmente compostas de uma série de unidades sensoriais que formam a camada de entrada, camadas intermediárias ocultas e uma camada de saída. O sinal de entrada é propagado na rede para frente e camada por camada (HAYKIN, 2008).

Com o uso das camadas intermediárias, as MLP atingem um poder computacional muito maior, se comparadas às redes de camada única. Esta

vantagem deve-se ao fato de que, ao contrário das redes de camada única, as MLP conseguem tratar dados de problemas que não são linearmente separáveis. Em teoria, com o uso de duas camadas intermediárias, as MLP podem ser usadas para implementar qualquer função, sendo ela linearmente separável ou não. A qualidade dos resultados da implementação da função irá depender do número de nodos que estão dispostos nas camadas intermediárias. A Figura 08 ilustra uma rede MLP com duas camadas intermediárias. (BRAGA, CARVALHO e LUDERMIR, 2011).

Figura 08 – Rede MLP típica com duas camadas intermediárias



Fonte: Adaptado pelo autor de Haykin (2008).

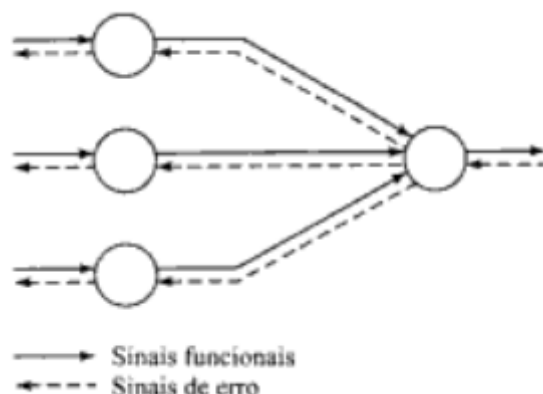
Conforme Braga, Carvalho e Ludermir (2011), nas redes com múltiplas camadas, cada uma delas possui responsabilidades e funcionamento distintos. O processamento que cada nodo realiza nestas camadas é definido pela combinação do processamento realizado nos nodos das camadas anteriores que estão interligados a ele. As funções implementadas nos nodos ficam cada vez mais complexas na medida em que se segue da primeira camada intermediária em direção à camada de saída. A divisão do espaço de decisão é definida por estas funções.

Segundo Braga, Carvalho e Ludermir (2011), às unidades intermediárias de uma MLP podem ser definidos como detectores de características. Estas unidades codificam os padrões de entrada que é então utilizado para a definição de saída da rede. Com a utilização de um número grande o suficiente de nodos intermediários,

é possível gerar representações para conjunto de padrões de entrada de qualquer natureza.

Em redes perceptron de múltiplas camadas, é possível identificar dois tipos de sinais, ilustrados na Figura 09 e descritos a seguir (HAYKIN, 2008).

Figura 09 – Fluxos de sinal de uma rede perceptron de múltiplas camadas



Fonte: Haykin (2008).

O sinal funcional consiste em um sinal de entrada ou estímulo que se propaga para a frente, neurônio em neurônio, até se tornar um sinal de saída. É chamado de sinal funcional pois assume-se que sejam realizadas funções na camada de saída, além deste sinal ser calculado como uma função com base em suas entradas e pesos em cada neurônio que ele passa (HAYKIN, 2008).

O sinal de erro é originado em um neurônio de saída da rede e se propaga camada por camada, em sentido para trás. É chamado de sinal de erro pois o cálculo estimativo do neurônio possui independentemente uma função que necessita de um erro.

Segundo Haykin (2008), através do treinamento de forma supervisionada, com o popular algoritmo de treinamento supervisionado chamado algoritmo de retropropagação de erro (*error-back propagation*), perceptrons de múltiplas camadas são aplicados para resolver uma grande quantidade de problemas difíceis. Este algoritmo é baseado na regra de aprendizagem por correção de erro.

2.4.4.2 *Backpropagation*

Redes perceptron de múltiplas camadas aprendem praticamente da mesma maneira que as redes perceptron de uma única camada. A grande diferença é que em redes de múltiplas camadas, cada neurônio possui seu peso relacionado com suas entradas, sendo assim, existe um número muito maior de pesos que precisam ser ajustados quando ocorre um erro no processo de treinamento. O método de retropropagação, internacionalmente conhecido como algoritmo de *backpropagation*, atribui responsabilidade nos diferentes pesos inferidos em cada neurônio, configurando um aprendizado por erro (COPPIN, 2013).

Backpropagation é um algoritmo supervisionado que usa pares de entradas e saídas desejadas para ajustar os pesos de rede por meio de operações de correção de erros. O treinamento acontece em duas fases, cada uma em um sentido diferente da rede, chamadas de fase *forward* e fase *backward* (BRAGA, CARVALHO e LUDERMIR, 2011).

Conforme Haykin (2008), na fase *forward*, um vetor de entrada é disposto aos nós sensoriais da rede e, então, seu efeito se propaga pela rede, camada a camada, até um conjunto de sinais de saída ser produzido como resposta real. Nesta fase, os pesos sinápticos são sempre fixos. Já na fase *backward*, os pesos sinápticos são ajustados com base em uma regra de correção de erros. A resposta real da rede é subtraída de uma resposta desejada (alvo), produzindo assim um sinal de erro. O sinal de erro é então propagado para trás da rede, na direção contrária das conexões sinápticas, ganhando assim o nome de retropropagação de erro (*error back-propagation*). Os pesos sinápticos são então ajustados para que a resposta real gerada pela rede fique mais próxima da saída desejada.

2.4.5 Redes Neurais Convolucionais

Redes neurais convolucionais (RNC) são um tipo de redes neurais especiais que usam uma topologia de grade. Como exemplos de dados em formato de grade pode-se citar dados temporais em série, que formam uma grade com uma

dimensão com amostras em tempos de intervalos regulares e também dados de imagens, que podem ser colocadas como grades de duas dimensões de pixels. As redes neurais convolucionais tem sido tremendamente bem-sucedidas em aplicações práticas. O nome "rede neural convolucional" implica que a rede faz uso de uma operação matemática chamada de convolução. Convolução é um tipo especializado de operação linear. Redes neurais convolucionais são, resumidamente, redes neurais que usam convolução e não uma matriz de multiplicação em pelo menos uma de suas camadas (GOODFELLOW et al., 2016).

As redes neurais convolucionais emergiram do estudo do córtex visual do cérebro e são usadas no reconhecimento de imagens desde a década de 1980. Esses estudos do córtex visual inspiraram a rede neural Neocognitron, introduzida em 1980, que evoluiu gradualmente para o que chamamos hoje de redes neurais convolucionais. Nos últimos anos, graças ao aumento do poder computacional, à quantidade de dados de treinamento disponíveis e as técnicas de treinamento de redes profundas, as CNNs conseguiram alcançar um desempenho sobre-humano em algumas tarefas visuais complexas (GÉRON, 2017).

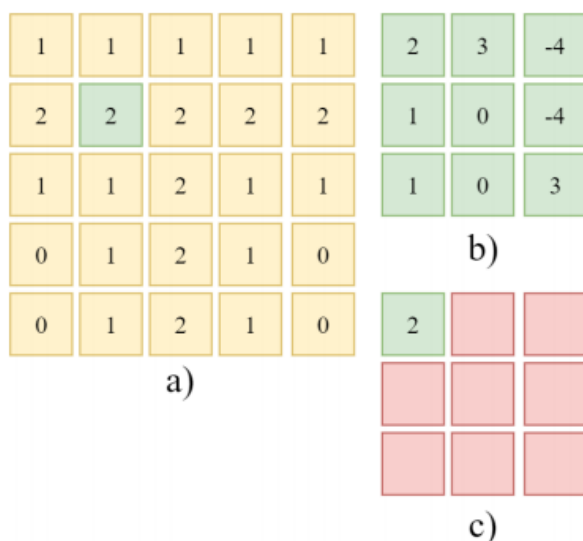
Para Ujjwal Karn (2016) uma RNC é composta basicamente por quatro tipos de operações: convolução, ReLU, *pooling* e classificação.

2.4.5.1 Convolução

O objetivo das camadas de convolução é aprender, eficientemente, características relevantes sobre a imagem. Para isto, é usada a técnica da conectividade local que também é utilizada pelo córtex cerebral, onde o campo receptivo do neurônio é chamado de filtro. O filtro é responsável por processar a imagem ou a camada anterior da rede, realizando operações de soma e produto dos valores em seu campo receptivo. Este processo, chamado de convolução, é realizado em todas as posições da entrada do neurônio (CERENTINI, 2018). A Figura 10 representa este processo: em b), o filtro de tamanho 3 x 3 percorre todos os pontos possíveis de a) resultando em c). Os valores de b) são ajustados pela

rede para resultar em mapas de características de maior relevância para as próximas camadas.

Figura 10 – Processo de convolução de uma RNC



Fonte: Cerentini (2018).

Para realizar o processo, filtro recebe uma matriz com 3 dimensões formada por altura, largura e profundidade. A largura e altura representam o tamanho do campo receptivo e a profundidade do número de canais da imagem ou camada anterior. Normalmente apenas o tamanho do campo receptivo passa por ajustes. O filtro utilizará todos os canais da mesma localidade para realizar as operações. Como o filtro é o mesmo para todos os neurônios que formam o campo visual, é possível detectar características na imagem independentemente de sua posição (KARPATHY, 2017).

Cerentini (2018) explica que a convolução realizada pelos filtros nas entradas da rede geram mapas de características. Os mapas de características guardam os locais onde o filtro detectou uma característica importante. A quantidade de diferentes filtros presentes na camada convolucional é quem dita o número destes mapas. Os neurônios que utilizam o mesmo filtro marcam também a intensidade no mesmo mapa, baseado na possibilidade de característica filtrada estar presente na sua região filtrada. Estes mesmos mapas são utilizados pelas próximas camadas da rede que realizam o mesmo processo, que ocorre até ao final da rede, onde representações próximas do objeto buscado estarão presentes.

Os dados resultantes da convolução são processados por uma função de ativação. A função de ativação mais utilizada neste caso é a de Unidade Linear Retificada (ULR), também conhecida por ReLU (*Rectified Linear Units*). Essa função substitui todos os valores negativos por zero, introduzindo não linearidade a rede, uma vez que a rede tem o propósito de resolver problemas não lineares (KARN, 2016).

Após uma camada convolucional, geralmente existe uma camada de *pooling*. O objetivo dessa camada é reduzir a dimensão espacial do volume de entrada, consequentemente diminuindo o custo computacional da rede e evitando *overfitting* (KARPATHY, 2017). Na operação de *pooling*, os valores pertencentes a uma região do mapa de características são substituídos por alguma métrica dessa região. A forma mais comum de *pooling* consiste em substituir os valores de uma região pelo valor máximo, como ilustra a Figura 11. Essa operação é conhecida como *max pooling* e é útil para eliminar valores desprezíveis, reduzindo a dimensão da representação dos dados e acelerando a computação necessária para as próximas camadas, além de criar uma invariância a pequenas mudanças e distorções locais. Outras funções populares de agrupamento incluem a média de uma vizinhança retangular, a norma L2 de uma vizinhança retangular ou uma média ponderada com base na distância do pixel central (GOODFELLOW et al., 2016).

Figura 11 – Aplicação de *max pooling* em imagem 4x4 utilizando um filtro 2x2

Matriz da imagem

2	1	3	1
1	0	1	4
0	6	9	5
7	1	4	1

Max pool

2	4
7	9

Fonte: Adaptado pelo autor de Walters (2019).

A última camada da rede tem a responsabilidade de classificar as entradas em uma das classes disponíveis. Normalmente esta camada é composta por neurônios inteiramente conectados, os mesmos vistos em redes neurais comuns,

que interpretam os últimos mapas de característica da rede. Essa interpretação ocorre devido aos neurônios desta camada estarem inteiramente conectados com todos os mapas de características da camada anterior, sendo capaz de misturar essas informações a fim de gerar uma visão geral da imagem.

Neste capítulo foram apresentadas revisões literárias sobre o olho humano, glaucoma, inteligência artificial e redes neurais. Iniciando com o olho humano, seguido de glaucoma foi possível entender como funciona o olho e os impactos causados pela doença de glaucoma. Após isso, foram então explicados assuntos referentes a inteligência artificial e então as redes neurais, tratando das redes neurais convolucionais, que foram utilizadas no presente trabalho.

3 TRABALHOS RELACIONADOS

Este capítulo descreve alguns trabalhos que utilizaram Aprendizado de Máquina por meio de RNAs para classificação e identificação de glaucoma em imagens de fundo do olho. O intuito foi obter conhecimento para auxiliar na estruturação do trabalho proposto e para que possa ser feito um comparativo com os resultados encontrados no presente estudo.

Raghavendra et al. (2018) propuseram uma ferramenta de diagnóstico assistido por computador para detecção precisa de glaucoma usando a técnica de aprendizado profundo. Uma rede neural convolucional de 18 camadas foi treinada efetivamente para extrair características importantes de imagens de fundo de olho. O processo de classificação utilizado foi de duas classes: normal e com glaucoma.

A ferramenta utilizada para implementar a RNC foi o *software* MATLAB. Para treinamento e validação da rede, foi utilizada uma base de imagens extraída do *Kasturba Medical College*, de Manipal, Índia. Para extração dos resultados foram realizadas repetições (iterações) nos experimentos cinquenta vezes e a média de todos os parâmetros de desempenho para toda iteração foram calculados. A precisão obtida foi de 98,13%, sensibilidade 98%, especificidade 98.30% e *positive predictive value* (PPV) 98.79%.

Sevastopolsky (2017) fez um estudo para analisar o uso de uma rede neural convolucional para realizar a segmentação do disco óptico e da escavação. A arquitetura de rede utilizada foi uma U-Net modificada.

As modificações feitas a arquitetura U-Net feitas por Sevastopolsky (2017) foram o uso de um menor número de filtros em todas as camadas convolucionais e a utilização do mesmo número de filtros em todas as camadas no caminho de contração.

A metodologia empregada foi uma divisão em duas etapas: um pré-processamento e após o treinamento com o uso da rede U-Net modificada para segmentação. No pré-processamento foi feito uma melhora no contraste das imagens, além de um recorte pela área de interesse. As imagens de fundo de olho utilizadas foram das bases de dados DRIONS-DB, DRISHTI-GS e RIM-ONE-R-3.

O método de Sevastopolsky (2017) utilizou a métrica *F1-score* para avaliar os resultados. Para a segmentação do disco óptico, obteve *F1-score* de 0,95 na base de dados RIM-ONE-R3 e 0,94 na base de dados DRIONS-DB. Para a segmentação da escavação, as métricas de *F1-score* obtidas foram 0,85 na base de dados RIM-ONE-R3 e 0,82 na base de dados DRISHTI-GS.

Lima (2019), desenvolveu sua pesquisa usando uma técnica chamada de otimização de hiperparâmetros para selecionar os melhores parâmetros de um algoritmo genético, visando escolher a arquitetura mais adequada de redes neurais convolucionais, através de técnicas evolutivas. O objetivo do trabalho de Lima (2019) foi classificar imagens de fundo em olho como glaucomatosas ou normais.

A metodologia do trabalho foi dividida em quatro principais etapas: pré-processamento das imagens; otimização dos hiperparâmetros de um AG; com o uso do AG, encontrar a melhor arquitetura de CNN; treinar e executar a CNN para classificação das imagens. A base de imagens utilizada foi a RIM-ONE, composta por 455 imagens de fundo do olho. Um pré-processamento foi aplicado às imagens, gerando um redimensionamento para o tamanho 96x96, no padrão RGB. Após o treinamento, os resultados da classificação apresentados pela RNC demonstraram uma acurácia de 96,63%, especificidade de 98,00%, sensibilidade de 94,87%, precisão de 97,37% e *F1-score* de 96,10%.

Cerentini (2018), utilizou um sistema de detecção de objetos chamado de YOLO9000, para detectar o nervo óptico, que é a região de interesse, além de uma

rede neural convolucional, para detectar a presença do glaucoma. Diferentes algoritmos classificadores foram usados a fim de verificar qual possui maior acurácia na classificação das imagens de fundo em olho como glaucomatosas ou normais.

As ferramentas utilizadas foram a linguagem de programação Python, o *framework* para aprendizado de máquina PyTorch, um *framework* para a detecção da região de interesse chamado de Darknet e uma biblioteca com nome de Augmentor para o aumento de dados. Imagens de fundo do olho da base de dados RIM-ONE e HRF foram usadas para treinamento e validação.

Os resultados obtidos pelo método de Cerentini (2018), utilizando a métrica da área da curva de característica de operação do receptor, foram 100% na base de dados HRF, 85,8% na base de dados RIM-ONE-R1-R2 e 94,4% em RIM-ONE-R3.

Yu et al. (2019) realizaram uma pesquisa utilizando uma rede neural convolucional, com arquitetura U-Net modificada e pré treinada utilizando ResNet, para desenvolver uma ferramenta capaz de segmentar e realizar cálculo de CDR em imagens de fundo de olho coloridas.

Para atingir seus objetivos, a metodologia da pesquisa empregada pelos autores teve duas etapas principais. As imagens de fundo do olho foram primeiramente pré-processadas, onde foi extraída a região de interesse, seu contraste aumentado e redimensionadas para 512 × 512 pixels. Então, com o uso de uma RNC com arquitetura modifica U-Net foi aplicada para realizar a segmentação do disco óptico e da escavação. A base de imagens RIGA foi utilizada para treinamento da rede, já as bases de imagens DRISHTI-GS e RIM-ONE V3 foram utilizadas para validação do modelo.

Com estas características, o trabalho de Yu et al. (2019) teve os resultados equivalentes com técnicas do estado da arte: *F1-score* médio do disco óptico de 97,38% e *F1-score* médio da escavação de 88,77% na base de imagens DRISHTI-GS; *F1-score* médio do disco óptico de 96,10% e *F1-score* médio da escavação de 84,45% na base de imagens RIM-ONE.

Os trabalhos relacionados, citados neste capítulo, comprovam que o uso de técnicas de inteligência artificial, com o uso de RNCs, no auxílio do diagnóstico do glaucoma, é uma importante ferramenta que tende a trazer benefícios para a área da saúde. O Quadro 02 apresenta um resumo dos trabalhos apresentados neste capítulo, alguns dos resultados neles atingidos assim como os resultados obtidos pelo autor no presente trabalho.

Quadro 02 – Resumo dos trabalhos relacionados e resultados obtidos

Autor	Base de imagens	Métricas e resultados
Raghavendra et al. (2018)	<i>Kasturba Medical College</i>	Acurácia: 98,13%; Especificidade: 98,30%; Sensitividade: 98%; <i>Positive predictive value (PPV)</i> : 98,79%.
Sevastopolsky (2017)	RIM-ONE-R3	<i>F1-score</i> médio do disco óptico: 0,95; <i>F1-score</i> médio da escavação: 0,82.
Sevastopolsky (2017)	DRISHTI-GS	<i>F1-score</i> médio da escavação: 0,85.
Sevastopolsky (2017)	DRIONS-DB	<i>F1-score</i> médio do disco óptico: 0,94.
Lima (2019)	RIM-ONE	Acurácia: 96,63%; Especificidade: 98,00%; sensibilidade: 94,87%; Precisão: 97,37%; <i>F1-score</i> : 96,10%.
Cerentini (2018)	HRF	Área da curva de característica de operação do receptor: 100%.
Cerentini (2018)	RIM-ONE-R1-R2	Área da curva de característica de operação do receptor: 85,8%.
Cerentini (2018)	RIM-ONE-R3	Área da curva de característica de operação do receptor: 94,4%.
Yu et al. (2019)	DRISHTI-GS	<i>F1-score</i> médio do disco óptico: 97,38%; <i>F1-score</i> médio da escavação: 88,77%.
Yu et al. (2019)	RIM-ONE	<i>F1-score</i> médio do disco óptico: 96,10%; <i>F1-score</i> médio da escavação: 84,45%.
Do autor (2020)	RIM-ONE-R3	<i>F1-score</i> médio do disco óptico: 0,9440; <i>F1-score</i> médio da escavação: 0,8350.

Fonte: Elaborado pelo autor (2020).

4 MATERIAIS E MÉTODOS

É imprescindível a escolha e a definição da metodologia para o processo de pesquisa científica, para que assim ela dite caminhos que devem ser percorridos para a realização com sucesso deste projeto. O método científico utilizado no presente trabalho será apresentado neste capítulo, juntamente às ferramentas e procedimento técnicos utilizados para idealização da proposta.

4.1 Classificação da pesquisa

O presente trabalho buscou avaliar se a rede de arquitetura U-Net é capaz de segmentar imagens de fundo do olho, conforme explicado anteriormente. Para alcançar este objetivo, pesquisas sobre os temas relacionados foram realizadas. Com o estudo desses temas será possível desenvolver a ferramenta que esteja dentro de padrões existentes, definindo assim, atributos para que este trabalho seja definido como pesquisa exploratória (SANTOS, 1999).

Segundo Marconi e Lakatos (2011), um trabalho de pesquisa exploratório é realizado em três etapas, o desenvolvimento de uma hipótese, a construção de um referencial teórico robusto para ampliar o conhecimento sobre os assuntos do estudo e, por último, a identificação de um fenômeno ou fato.

Com o intuito de enriquecer o trabalho e para que este estivesse em conformidade com padrões já estabelecidos, foi necessário o levantamento de

informações acerca do tema estudado, levando assim a necessidade do uso do método de pesquisas bibliográficas. Segundo Gil (2002), neste método de pesquisa faz-se uso de referências em bibliografias a respeito do tema estudado, para assim poder realizar uma melhor fundamentação do trabalho científico.

Como forma de avaliação da abordagem proposta, a presente pesquisa foi classificada como qualitativa. Ramos (2009) explica que a abordagem qualitativa é utilizada nas situações de estudo em que, por conta de sua complexidade, torna-se difícil de aplicar uma quantificação aos resultados obtidos. Na inexistência de uma hipótese a ser comprovada, o seu uso é apropriado para o entendimento do fenômeno estudado, definindo assim uma conclusão indutiva sobre o assunto.

4.2 Tecnologias

Nesta seção são elencadas as principais ferramentas utilizadas no desenvolvimento do presente trabalho, incluindo equipamentos, linguagens de programação e bancos de imagens.

4.2.1 Equipamentos

Para realizar os experimentos deste trabalho foi utilizado um computador pessoal com as seguintes configurações:

- Processador: AMD Ryzen 7 2700X Com 8 núcleos de 3.2GHz;
- Memória: HyperX Fury 8GB 2933 MHz DDR4;
- Placa de vídeo: NVIDIA GeForce GTX 1060 com 6GB de memória;
- Armazenamento: SSD Corsair Force MP300 com 248 GB;
- Sistema operacional: Windows 10.

4.2.2 Bancos de imagens

Nesta seção, está descrito o *dataset* de imagens utilizadas para a construção deste projeto. O *dataset* de imagens é composto por imagens de fundo de olho originais, captadas por um equipamento de rinoplastia e analisadas por profissionais, além de outras imagens com informações de localização do disco óptico e da escavação.

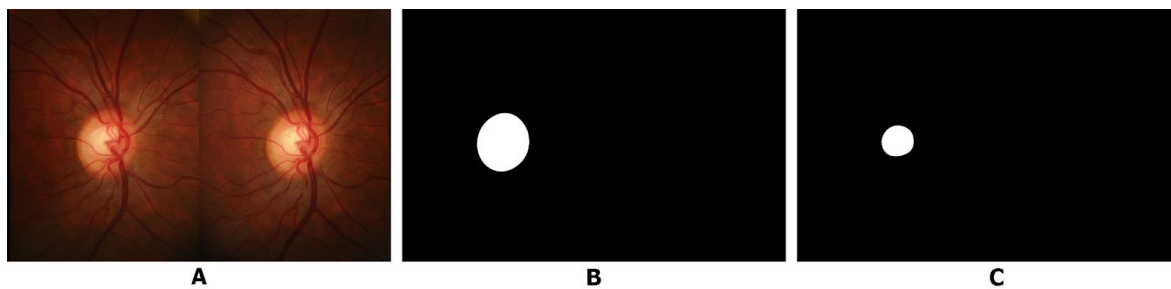
4.2.2.1 RIM-ONE R3

A base de dados RIM-ONE R3 é a terceira versão da RIM-ONE, (*Retinal Image database for Optic Nerve Evaluation*). Ela contém um total de 159 imagens de fundo de olho, sendo 74 imagens de glaucoma ou suspeita e 85 imagens saudáveis. As imagens foram capturadas usando o Nidek AFC-210 com um corpo de uma Canon EOS 5D Mark II de 21,1 megapixels (FUMERO et al., 2011).

Além das imagens de fundo de olho, RIM-ONE R3 contém as máscaras de segmentação *ground-truth* (valor de referência), para segmentação de disco óptico e escavação realizados por dois especialistas em oftalmologia do *Hospital Universitario de Canaria*, de Ilhas Canárias na Espanha. Compõe também uma outra máscara representando a média dos pixels da segmentação realizada pelos especialistas. Este valor de referência da segmentação foi realizado com o uso do *software* DCSeg.

A Figura 12, demonstra exemplo de uma imagem da base de imagens RIM-ONE V3 de um paciente diagnosticado como normal (A), além de sua respectiva máscara média do disco óptico (B) e máscara média da escavação (C).

Figura 12 – Imagem original RIM-ONE R3 e suas máscaras



Fonte: Adaptado pelo autor de Fumero et al. (2011).

4.2.3 Python

Python é uma linguagem de programação orientada a objetos interpretada e interativa, combina um poder remarcável e uma sintaxe bem clara. Pode ser estendida em C ou C++. Como é uma multiplataforma, é possível executá-la em Mac, Unix e Windows (PYTHON, 2018).

Python foi desenvolvido inicialmente por Guido Van Rossum no final dos anos 80. Enquanto trabalhava no projeto Amoeba, na empresa CWI, tinha liberdade para fazer seus subprojetos dentro do escopo do que estava trabalhando. Decidiu então utilizar o aprendizado que teve em um projeto de uma linguagem de programação na qual havia trabalhado anteriormente, chamada de ABC, para criar um design diferente para esta linguagem, fazendo com que esta tivesse as melhores propriedades da ABC, porém, sem seus problemas. Assim se deu início ao python (VENNERS, 2003).

4.2.3.1 Bibliotecas python

As principais bibliotecas do python utilizados no desenvolvimento deste trabalho, desde os algoritmos de tratamento de imagens até a rede neural estão listados a seguir:

- **Python Imaging Library (PIL):** Biblioteca que adiciona suporte a operações com imagem de diferentes formatos como abrir, manipular e salvar (CLARK, 2020);
- **Scikit-image:** Biblioteca que possui uma coleção de algoritmos para o processamento de imagens (VAN DER WALT et. al., 2014);
- **Scikit-learn:** Biblioteca que dispõe de diversos algoritmos de aprendizado de máquina (PEDREGOSA et. al., 2011);
- **Numpy:** Biblioteca que adiciona suporte a grandes vetores, vetores multi-dimensionais e matrizes além de uma grande quantidade de funções matemáticas para são operadas nestes vetores (VAN DER WALT et. al., 2011);
- **OpenCv:** Biblioteca para visão de computador, aprendizado de máquina, processamento de imagem e para um grande número de outras operações de processamento em tempo real (OPENCV, 2019).

Além das listadas anteriormente, outras duas bibliotecas extremamente importantes para o desenvolvimento deste trabalho foram TensorFlow e Keras, descritas nas seções a seguir.

4.2.3.1.1 TensorFlow

O TensorFlow é uma poderosa biblioteca de *software* para computação numérica, particularmente adequado e refinado para uso com aprendizado de máquina de grande escala. Possui capacidade de treinar redes neurais com milhões de parâmetros em uma estrutura de treinamento composta por bilhões de instâncias com milhões de funções cada (GÉRON, 2017).

Os responsáveis pelo desenvolvimento do TensorFlow fazem parte de um departamento do Google chamado de Google Brain. Este departamento é responsável pela criação e pesquisas na área de IA com foco em redes neurais profundas. Esta equipe promove pesquisas que aprimoram o estado da arte de algoritmos de IA e estas novas técnicas são aplicadas em campo. Atuam em áreas

de grande impacto social como, saúde, segurança, energia, transporte, manufatura e entretenimento (GOOGLE AI, 2019).

Anos de experiência dos desenvolvedores do antigo projeto chamado de *DistBelief* é que deu origem ao TensorFlow, gerando uma ferramenta simplificada e genérica para permitir que pesquisadores explorem uma ampla variedade de ideias com facilidade. O TensorFlow suporta treinamento e inferência em larga escala, executando eficientemente centenas de servidores (habilitados para GPU) para rápido treinamento e executa modelos treinados para inferência em várias plataformas, desde grandes *clusters* distribuídos até em dispositivos móveis. Ao mesmo tempo, é flexível o suficiente para apoiar a experimentação e pesquisa de novos modelos de aprendizado de máquina e otimizações no nível do sistema.

O TensorFlow é utilizado em vários produtos em larga escala do Google como, por exemplo, Cloud Speech, Google Photos e Google Search (GÉRON, 2017).

4.2.3.1.2 Keras

Keras é uma *Application Programming Interface* (API) para manipulação de redes neurais em alto nível. Escrita na linguagem de programação Python, ela é capaz de trabalhar juntamente com outras ferramentas como TensorFlow, CNTK ou Theano. O seu foco de desenvolvimento foi permitir uma experimentação rápida, fazendo com que ideias se tornassem resultados. Foi inicialmente desenvolvido como parte de uma pesquisa para o projeto *Open-ended Neuro-Electronic Intelligent Robot Operating System* (ONEIROS) (CHOLLET et al., 2015).

A biblioteca Keras permite execução em CPU e GPU, possui suporte para redes neurais convolucionais, redes neurais recorrentes e combinações das duas, além de possuir uma sintaxe clara e objetiva. Foi desenvolvida para ter uma interface de uso amigável e facilitar o trabalho dos desenvolvedores (CHOLLET et al., 2015).

4.2.4 Arquitetura de rede neural U-Net

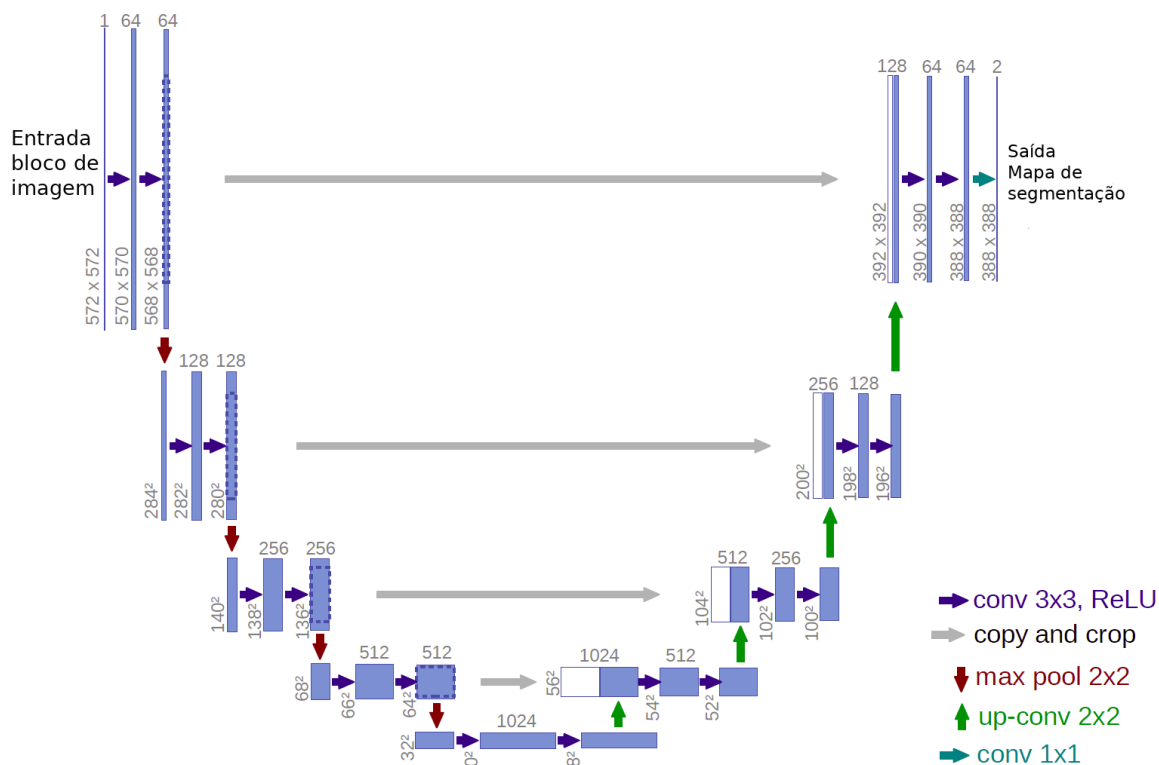
U-Net é uma arquitetura de rede neural convolucional em formato de U originalmente desenvolvida para segmentação de imagens biomédicas (RONNEBERGER et al., 2015). A arquitetura desta rede, ilustrada na Figura 13 é composta por um caminho de contratação (lado esquerdo) e um caminho expansivo (lado direito).

O caminho da contratação segue a arquitetura padrão de uma rede convolucional. Este consiste na aplicação de duas convoluções 3x3, cada uma seguida de uma unidade linear retificada (ReLU) e uma operação de *max pooling* 2x2 com *stride* 2 para *downsampling*, que é a redução do tamanho da imagem. Em cada etapa de *downsampling*, dobramos o número de canais de características.

Cada etapa do caminho expansivo consiste em uma operação de *upsampling* do mapa de características, que é o aumento do tamanho da imagem, seguida por uma convolução 2x2 ("convolução ascendente") que reduz pela metade o número de canais de recursos, uma concatenação com o mapa de características cortadas correspondente ao caminho de contratação e duas convoluções 3x3, cada uma seguida por uma ReLU. O corte é necessário devido à perda de pixels da borda em todas as convoluções.

Na camada final, uma convolução 1x1 é usada para mapear cada vetor de característica de 64 componentes para o número desejado de classes. No total, a rede possui 23 camadas convolucionais.

Figura 13 – Arquitetura de uma rede U-Net

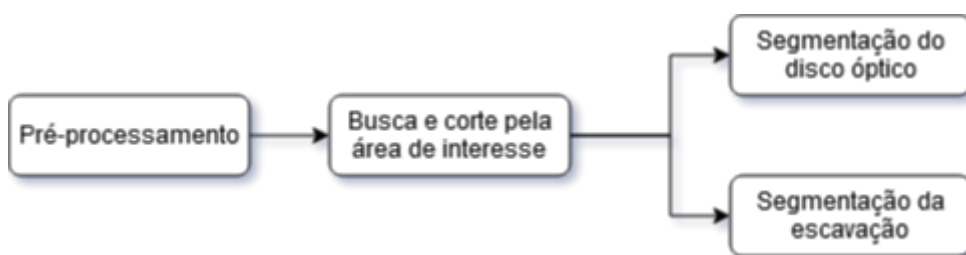


Fonte: Adaptado pelo autor de Ronneberger et al. (2015).

4.3 Desenvolvimento

O desenvolvimento da parte técnica deste trabalho foi realizado em etapas. Primeiramente uma etapa de pré-processamento foi realizada na base das imagens para adequar e padronizar os dados. Após, foi aplicado um método para encontrar a área de interesse nas imagens com o uso de uma rede neural convolucional de arquitetura U-Net. Por fim, outras duas etapas independentes com o uso da mesma arquitetura de RNC, sendo uma delas a segmentação do disco óptico e a outra a segmentação da escavação. A descrição completa e os resultados obtidos pelas duas últimas etapas são informados no capítulo 6. A Figura 14 ilustra o fluxo de atividades empregadas no desenvolvimento deste trabalho.

Figura 14 – Fluxo das atividades



Fonte: Elaborado pelo autor (2020).

4.3.1 Pré-processamento

A etapa de pré-processamento é a etapa onde as imagens são preparadas de uma forma que possam alimentar a rede neural. Esta etapa é importante para o projeto pois ajuda a resumir os dados que serão utilizados na rede neural, além de fornecer dados de uma maneira mais clara para a rede neural, impactando assim nos seus resultados obtidos.

A base de dados RIM-ONE R3, como descrita anteriormente, contém imagens de dimensões 2144px por 1424px. Como cada imagem é composta por 2 imagens de fundo de olho referentes ao mesmo olho como mostra a imagem 12, foi realizado um corte na imagem, selecionando apenas a imagem da esquerda. Com o corte, as dimensões da imagem ficaram de 1072px por 1424px. Para gerar uma simplificação nas dimensões, foi adicionado uma borda preta nos lados esquerdo e direito da imagem para tornar as dimensões da imagem simétricas. Após isso, seu tamanho foi reduzido para 512px por 512px.

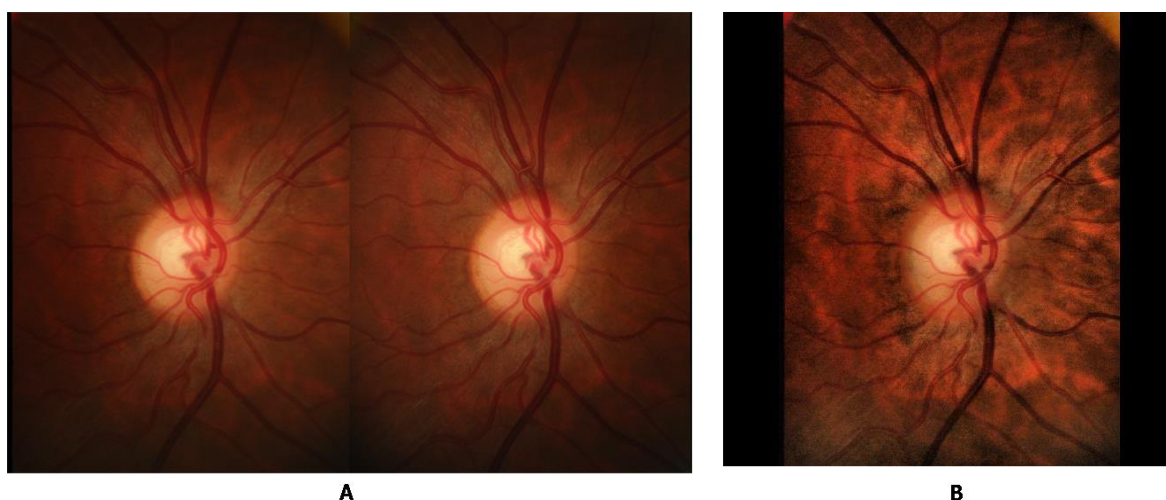
É possível identificar visualmente que as imagens do *dataset* estão escuras, com pouco contraste. Um processamento adicional poderia ser feito para melhorar a distinção dos elementos presentes na imagem. Testes realizados pelo autor, precedente a este trabalho, comprovaram que, com técnicas de melhoria nestas características, a rede neural obtém melhores resultados.

Setiawan et al. (2013) verificaram que utilizando a técnica de *Contrast-Limited Adaptive Histogram Equalization* (CLAHE) em imagens de fundo de olho foi

possível melhorar a qualidade da mesma. CLAHE é uma técnica de processamento de imagens que visa aperfeiçoar o contraste através da equalização de histograma, distribuindo os valores de luminosidade por toda a imagem. Com base nestas observações, foi aplicada a técnica de CLAHE em todas as imagens de fundo de olho do *dataset*.

A Figura 15 demonstra uma imagem resultante dos procedimentos descritos nesta seção (B) em comparação com uma imagem original do *dataset* (A).

Figura 15 – Imagem de fundo de olho original e após pré-processamento



Fonte: Adaptado pelo autor de Fumero et al. (2011).

Estas mesmas etapas de corte e redimensionamento precisaram ser aplicadas nas imagens das máscaras do disco óptico e da escavação que estão presentes na base de imagens, para que a classificação dos pixels não seja perdida.

Após as etapas de pré-processamento mencionadas acima, as 159 imagens e máscaras do *dataset* foram divididas em 3 conjuntos: um conjunto de treinamento com pouco mais de 76% (123) das imagens; um conjunto de 12% (18) das imagens para validação; e o último também com 12% (18) das imagens, para testes após o processo de treinamento.

Como resultado da divisão do *dataset* pelo autor, cada um dos conjuntos de teste e validação ficou composto por 6 imagens de fundo do olho para um dos diferentes diagnósticos: glaucoma, suspeito e normal. Já o conjunto de treinamento

ficou com 27 imagens de pacientes considerados pelos especialistas como glaucoma, 73 consideradas como normal e 23 suspeitos.

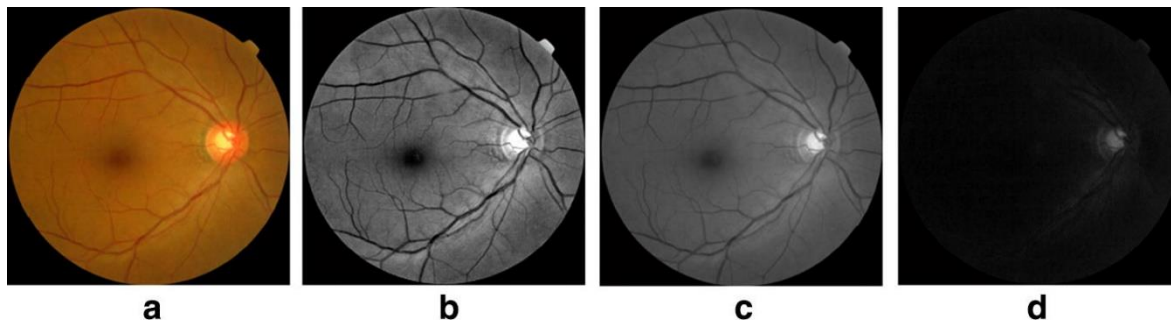
4.3.1.1 Utilização do canal G

Os componentes que formam uma imagem colorida, chamada de imagem RGB, são 3 camadas (canais) de cor onde cada canal é formado por apenas uma cor. Estes três canais são o canal de cor *Red* (R) vermelho, *Green* (G) verde e *Blue* (B) azul. Estes três canais de diferentes cores podem ser misturados, gerando assim novas cores. É a partir destes três canais de cor que as televisões, monitores e celulares representam as imagens.

A separação de uma imagem em seus diferentes canais de cor pode revelar características de maneiras diferentes, podendo assim ser utilizada para melhorar o contraste para algo específico presente na imagem.

Como mostra o estudo de Yang et al. (2018), o canal G da imagem de fundo de olho é o canal onde é possível visualmente identificar o maior nível de contraste para o nervo óptico. A da Figura 16 ilustra esta constatação onde em “A” é representado uma imagem de fundo de olho normal; em “B” a imagem do canal *Red* (R); em “C” a imagem do canal *Green* (G); e em “D” a imagem do canal *Blue* (B).

Figura 16 – Canais em uma imagem RGB do fundo do olho



Fonte: Yang et al. (2018).

Levando em consideração esta constatação, apenas o canal G, das imagens de fundo de olho foi utilizado durante este estudo. A extração do canal G é feita em tempo real durante o processo de treinamento, validação e testes deste trabalho.

4.3.1.2 Aumento dos dados

Como o número de exemplos (imagens para treinamento) de 123 imagens é considerado pequeno, foi necessário utilizar técnicas para aumento dos dados (*data augmentation*). Como explicam Ronneberger et al. (2015), o forte uso de técnicas de aumento dos dados é essencial para ensinar a rede neural propriedades de invariância e robustez quando o número de exemplos for pequeno.

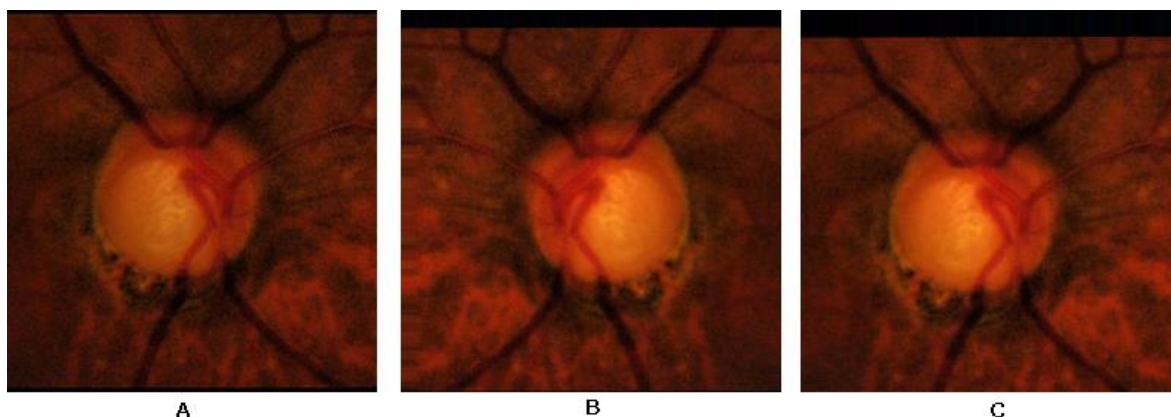
Para realizar o aumento dos dados foi utilizado uma classe python fornecida pelo Keras chamada de *ImageDataGenerator* que gera automaticamente novos exemplos com transformações de, por exemplo, zoom, inversão vertical e horizontal. Estas transformações foram realizadas de forma aleatória com base nos parâmetros informados, podendo assim criar um número enorme de possibilidades de exemplos.

Neste estudo, as transformações utilizadas no aumento dos dados foram:

- *rotation_range*: faixa de graus para rotações aleatórias, com valor de 0,2;
- *width_shift_range*: deslocamento aleatório horizontal das imagens, com valor de 0,05;
- *height_shift_range*: deslocamento aleatório vertical das imagens, com valor de 0,05;
- *shear_range*: ângulo de cisalhamento no sentido anti-horário, medido em radianos, com valor de 0,05;
- *zoom_range*: intervalo de zoom aleatório, com valor de 0,05;
- *horizontal_flip*: inverte aleatoriamente as entradas horizontalmente.

A Figura 17 ilustra dois exemplos resultantes do procedimento de aumento de dados. Em “A”, a imagem de fundo de olho do *dataset* após o pré-processamento. Em “B”, um exemplo de uma imagem aumentada, passando por um procedimento de inversão horizontal. Em “C”, outro exemplo de imagem aumentada onde o recebeu zoom e um deslocamento vertical para baixo. “B” e “C” foram geradas automaticamente pela classe *ImageDataGenerator* em tempo real durante procedimento de treinamento.

Figura 17 – Exemplos do resultado de aumento dos dados



Fonte: Adaptado pelo autor de Fumero et al. (2011).

Este procedimento de aumento dos dados foi aplicado ao conjunto de imagens de treino, sendo realizado em tempo real durante o processo de treinamento. Todos os procedimentos de treinamento realizados no presente estudo fizeram uso desta técnica de aumento de dados.

4.3.2 Arquitetura da rede neural

A arquitetura de rede neural utilizada neste trabalho foi a U-Net, desenvolvida por Ronneberger et al. (2015) e explicada na seção 4.2.4 deste documento.

Uma camada de convolução 1x1 foi incluída no modelo original de Ronneberger et al. (2015). Utilizando como função de ativação a função sigmóide, esta camada foi adicionada como camada da saída.

A função sigmóide foi escolhida para a camada de saída devido à natureza do problema que a rede neural foi construída para resolver ser uma análise probabilística: a probabilidade do pixel em questão, ser um pixel de interesse ou não. Sua ativação converte as entradas de qualquer valor negativo a qualquer valor positivo, resultando em um valor entre 0 e 1.

Para realizar a implementação em Python, nas camadas de convolução, a classe do Keras utilizada foi a `Conv2D`. Para *downsampling*, `Dropout`. Nas camadas de *max-pooling*, `MaxPooling2D`. Para *upsampling*, `UpSampling2D`. Para concatenação, foi utilizado a função `concatenate`.

Como o objetivo da rede neural é uma classificação binária, prever se um pixel é ou não é uma parte específica do olho, foi necessário utilizar uma função de perda (*loss function*) binária, neste caso a função *binary-crossentropy*. O otimizador (*optimizer*) utilizado foi o Adam com taxa de aprendizado (*learning-rate*) de 0,0001. Com estas características, a rede neural desenvolvida ficou composta de 31.031.685 parâmetros treináveis.

A rede recebe como parâmetro um vetor de dimensões 256 por 256 elementos. Como saída, retorna um vetor de mesmas dimensões contendo valores entre 0 e 1 representando a probabilidade de ser um pixel de interesse ou não. Como o objetivo é prever se o pixel é ou não de interesse, estes valores entre 0 e 1 são convertidos para binário: quando forem maiores que 0,5 são considerados verdadeiro, caso contrário, falso.

4.3.3 Configuração do treinamento

O procedimento de treinamento é o processo onde a rede neural aprende a resolver um problema, neste caso, a localizar pixels. Neste trabalho, três procedimentos distintos de treinamento foram realizados, sendo eles: treinamento para localização de pixels da área de interesse; treinamento para localização de pixels do disco óptico; e treinamento para localização de pixels da escavação. Cada

um dos procedimentos de treinamento possui particularidades individuais nos parâmetros que serão explicadas no decorrer deste documento.

No treinamento a questão que a rede deve aprender é a seguinte: onde estão localizados os pixels com tais características? Ou melhor, qual a probabilidade do pixel em questão ser um pixel com a característica desejada? Para ensinar a rede a resolver esta questão, ela recebe exemplos. Os exemplos são as imagens de fundo do olho e suas respectivas máscaras. As máscaras são indicativos binários da localização de interesse: 1 para os pixels que nos interessam e 0 para os que não nos interessam.

Python disponibiliza *generators*, que são agrupadores para dados que permitem a iteração de itens invocando para a memória os dados adicionados a ele apenas no momento que serão utilizados, reduzindo assim a utilização de memória. Dois *generators* são utilizados para agrupar as imagens que foram utilizadas no Keras durante a etapa de treinamento. Um *generator* contendo as imagens e máscaras do conjunto de imagens de teste e outro contendo as imagens do conjunto de validação. Estes *generators* contêm os dados criados na etapa de pré-processamento, com dados aumentados e com transformação para uso apenas do canal G das imagens, sendo os dois últimos gerados em tempo real durante o procedimento de treinamento.

Os treinamentos foram realizados com o uso da função do Keras chamada *fit_generator*. Nela são informados os parâmetros: *generators* de teste e validação, a quantidade de épocas (*epoch*), quantidade de etapas (*steps*) por *epoch*, *batch size* e *callbacks*. Cada um dos procedimentos de treinamento deste trabalho utilizou diferentes valores para estes parâmetros, eles serão descritos no decorrer deste documento.

Callbacks são funções executadas após o término de cada época. Uma época (*epoch*) refere-se a um ciclo completo de treino em uma rede neural, quando o conjunto inteiro de dados é analisado pela rede. Neste trabalho, duas *callbacks* foram utilizadas.

A primeira *callback* utilizada foi a da classe *ModelCheckpoint* do Keras. Ela foi responsável por armazenar em disco os pesos (*weights*) dos parâmetros treinados na rede neural, sendo configurada para salvar apenas os melhores resultados, estes quando a métrica *F1-score* máxima for atingida.

A segunda *callback* foi *EarlyStopping*. Ela foi responsável por interromper o treinamento após o mesmo não apresentar melhora na métrica *F1-score* máxima após o término de um número determinado de épocas. Este número de épocas é um dos parâmetros que será aprofundado no decorrer do documento.

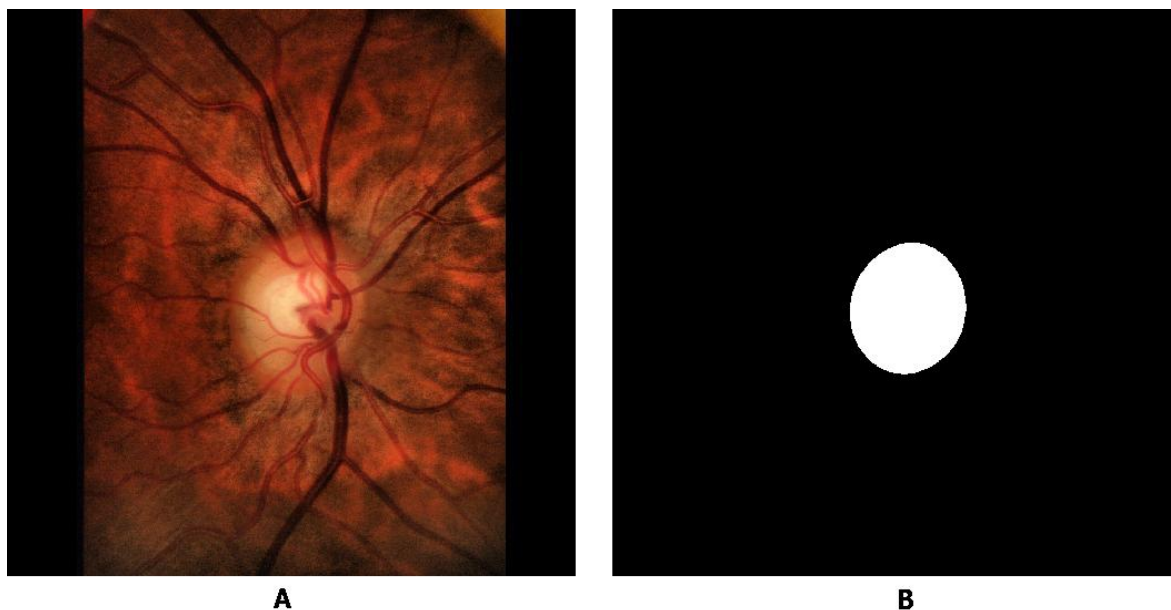
4.3.4 Busca automatizada da área de interesse

É possível verificar que, na Figura 15, a área de interesse, que é o nervo óptico em si, fica localizado bem no centro, cobrindo apenas uma pequena porcentagem do total de pixels da imagem. Desta forma, foi desenvolvido um método para cortar a imagem mais próxima a área de interesse, de uma forma em que a área de interesse cubra uma porcentagem maior.

Utilizando as imagens do disco óptico preparadas na etapa de pré-processamento e suas máscaras, a rede neural de arquitetura U-Net foi treinada para identificar a área de interesse. O treinamento para a localização da área de interesse utiliza as mesmas máscaras do treinamento para a segmentação do disco óptico. Após a localização, é adicionando uma margem adicional de pixels, resultando assim na área de interesse.

A Figura 18 representa um exemplo utilizado no procedimento de treinamento para automatização da busca da área de interesse. Em “A” uma imagem do fundo do olho e, em “B” sua respectiva máscara binária. Em branco são os pixels que fazem parte do disco óptico e em preto, os que não fazem parte.

Figura 18 – Imagem do fundo do olho e máscara do disco óptico



Fonte: Adaptado pelo autor de Fumero et al. (2011).

A quantidade de etapas por época utilizada neste procedimento de treinamento com a função *fit_generator* foi de 1.000 e a de épocas, 20. O *batch-size* foi de 2. Foram utilizadas as duas *callbacks* descritas na seção de treinamento, a *ModelCheckpoint* para armazenar os resultados e *EarlyStopping* configurada para aguardar por 8 épocas pela melhora dos resultados, caso contrário interromper o treinamento.

Após a época 9, onde a rede previu com métrica *F1-score* de 0,8817, não houve mais melhora na métrica. Sendo assim, o treinamento foi interrompido e os pesos encontrados na época 9 escolhido para uso nas previsões.

Com a rede neural treinada para localizar o disco óptico, foi então passado por ela todas as imagens de todos os conjuntos de treinamento, validação e testes para que, como saída, ela retornasse a previsão das coordenadas do disco óptico. Estas coordenadas foram armazenadas em arquivo.

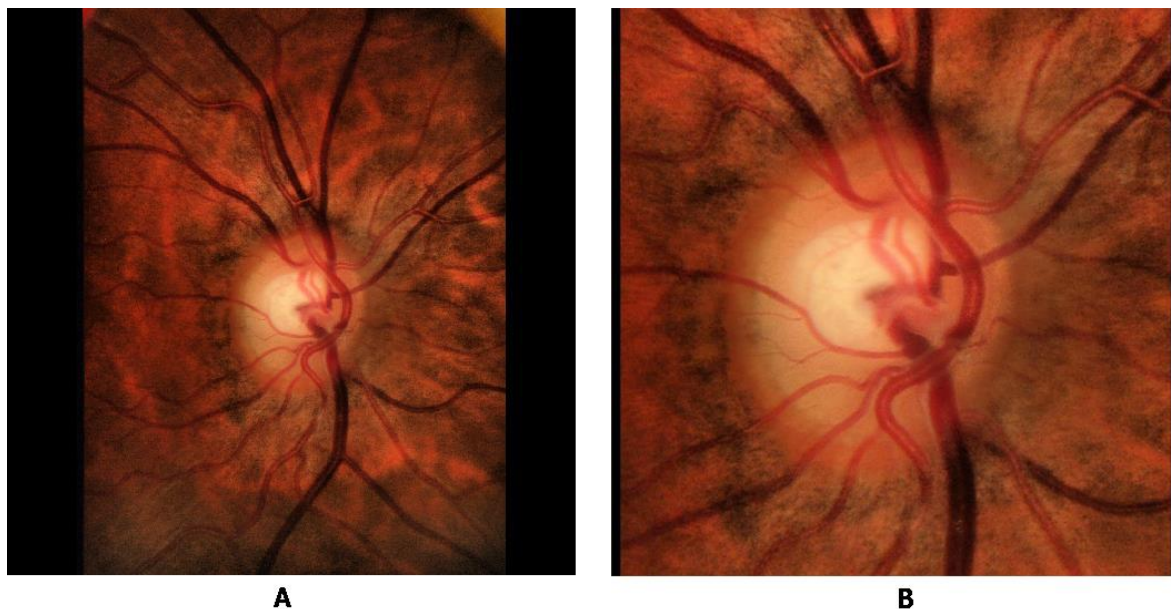
Com as coordenadas, foi possível realizar um corte mais específico na área de interesse. Para realizar o corte, foram adicionados às coordenadas 150 pixels no topo, na direita, na base e na esquerda, servido assim como margem para a imagem. Estas novas coordenadas foram extraídas das imagens originais

presentes no *dataset* e não das imagens geradas na etapa de pré-processamento devido às imagens originais estarem com uma resolução maior do que as imagens pré-processadas. Este procedimento gerou imagens com um corte mais preciso, porém, removeu as características geradas na etapa de pré-processamento.

Como o corte na área de interesse resultou na perda de características, o procedimento de pré-processamento foi realizado novamente nestas imagens resultantes. Nas imagens e máscaras resultantes, novamente foi realizado o procedimento para deixar a imagem com dimensões simétricas, evitando a perda de relação de aspecto. Após, foram redimensionadas para 512px por 512px, adicionando assim mais uma vez todas as características criadas no pré-processamento anterior, agora com um corte mais preciso na área de interesse.

A Figura 19 demonstra uma comparação de uma imagem resultante do procedimento de corte pela área de interesse “B” em comparação com o do pré-processamento aplicado anteriormente “A”.

Figura 19 – Comparação após corte da área de interesse com corte original



Fonte: Adaptado pelo autor de Fumero et al. (2011).

Os mesmos procedimentos de corte foram realizados também para as imagens e máscaras de disco óptico, máscaras da escavação de ambos os conjuntos de imagens para que não ocorresse perda da classificação dos pixels.

As etapas de segmentação do disco óptico e segmentação da escavação serão descritas no capítulo 5, juntamente com seus resultados para que se possa ter uma leitura e entendimento completo, desde sua implementação até seus resultados.

5 RESULTADOS E DISCUSSÕES

Este capítulo tem o propósito de apresentar os resultados da segmentação em imagens de fundo do olho das atividades executadas durante o desenvolvimento deste trabalho.

Os experimentos foram aplicados com dados que a rede neural ainda não fora exposta: o conjunto de testes do *dataset* que foi separado antes dos processos de treinamento da rede. Como descrito na seção de pré-processamento, o conjunto de testes é composto por 18 imagens, sendo elas 6 de pacientes diagnosticados com glaucoma, 6 suspeitos e 6 normais, originadas do *dataset* RIM-ONE R3.

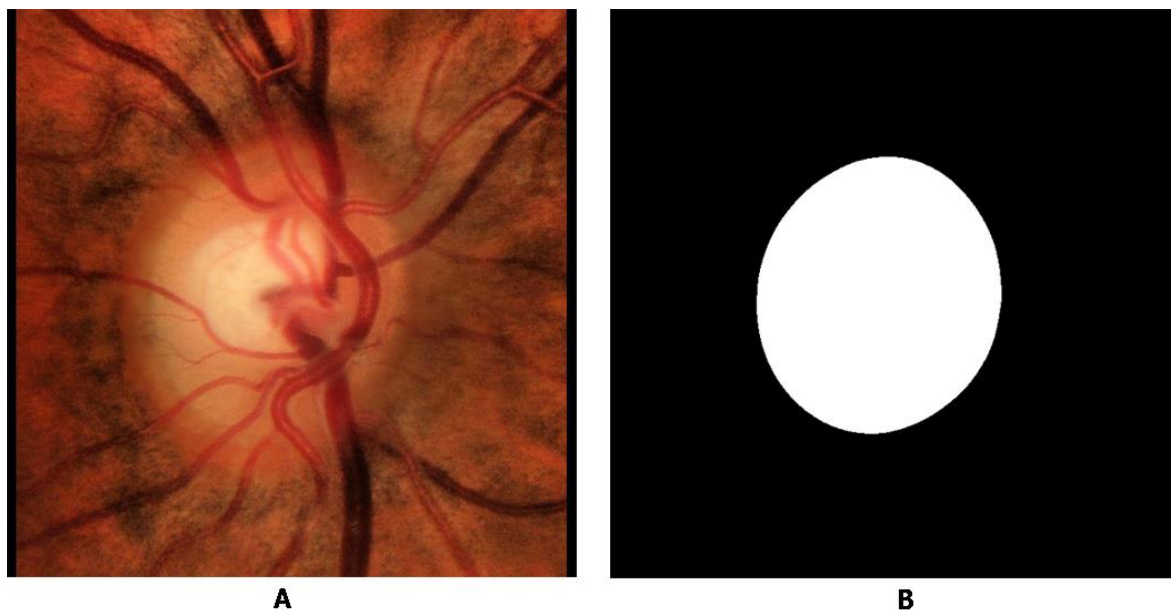
Para avaliação, foi utilizada a métrica *F1-score*, que é a média harmônica das métricas de precisão (*precision*) e revocação (*recall*). Uma boa métrica de *F1-score* é próxima de 1,00 e significa que há um pequeno número de casos falso positivos (FP) e pequeno número de falso negativos (FN). É comum o seu uso para avaliação em algoritmos de segmentação de imagens, sendo utilizada como métrica de avaliação nos trabalhos relacionados de Sevastopolsky (2017) e Yu et al. (2019).

5.1 Segmentação do disco óptico

Nesta etapa é onde a rede neural aprende a realizar a segmentação do disco óptico. Os exemplos que são utilizados neste treinamento são os resultantes do

procedimento de corte pela área de interesse. A Figura 20 representa um exemplo utilizado para treinamento da rede neural, em “A” uma imagem do fundo do olho e, em “B” sua respectiva máscara binária. Em branco são os pixels que fazem parte do disco óptico e em preto, os que não fazem parte.

Figura 20 – Imagem do fundo do olho cortada na área de interesse e máscara do disco óptico



Fonte: Adaptado pelo autor de Fumero et al. (2011).

A quantidade de etapas por época utilizada neste treinamento foi de 1.000 e a de épocas, 50. O *batch-size* foi de 2. Foram utilizadas as *callbacks ModelCheckpoint* para armazenar os resultados e *EarlyStopping* configurada para interromper o treinamento caso não ocorresse melhora dos resultados após 15 épocas.

A segmentação do disco óptico é a tarefa com que a rede tem maior facilidade de realizar devido a visível distinção entre o disco óptico e os demais componentes da imagem de fundo do olho.

O Quadro 03 apresenta as métricas obtidas no conjunto de dados de testes desta etapa. Cada linha representa uma imagem e a métrica obtida. Na coluna da esquerda fica o nome da imagem informando o diagnóstico do paciente presente do *dataset* e na coluna da direita, o seu valor de *F1-score*. Quanto mais próximo de

1,00 o *F1-score* estiver, melhor o desempenho da rede. Na última linha é apresentada a média aritmética dos valores encontrados.

Quadro 03 – Resultados da segmentação de disco óptico na base de dados RIM-ONE R3

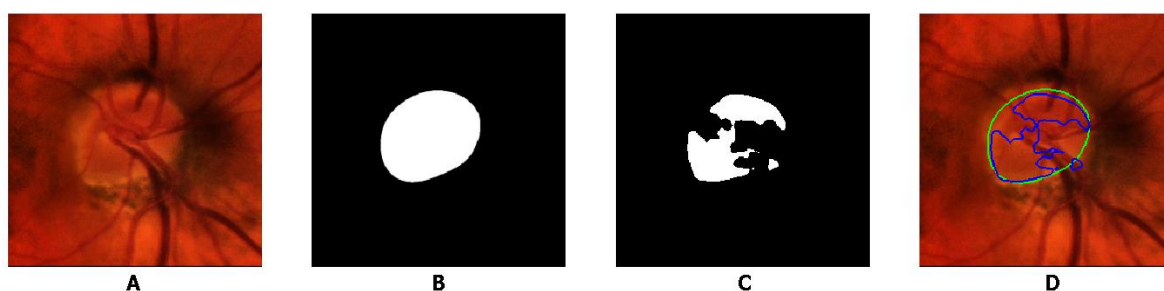
Nome da imagem	<i>F1-score</i>
Glaucoma 1	0,9300
Glaucoma 2	0,9816
Glaucoma 3	0,9453
Glaucoma 4	0,9615
Glaucoma 5	0,6818
Glaucoma 6	0,9700
Normal 1	0,9723
Normal 2	0,9624
Normal 3	0,9470
Normal 4	0,9366
Normal 5	0,9578
Normal 6	0,9549
Suspeito 1	0,9262
Suspeito 2	0,9793
Suspeito 3	0,9613
Suspeito 4	0,9716
Suspeito 5	0,9766
Suspeito 6	0,9757
Média	0,9440

Fonte: Elaborado pelo autor (2020).

Com exceção de uma imagem, os resultados de *F1-score* encontrados variam de 0,9262 chegando até 0,9816. O pior resultado encontrado foi em uma imagem de um paciente diagnosticado com glaucoma, Glaucoma 5, ilustrado na Figura 21, com *F1-score* de 0,6818. O resultado desta imagem ficou bem distante do segundo pior resultado, imagem do Suspeito 1, com *F1-score* de 0,9262. Já o melhor resultado foi da imagem Glaucoma 2, com *F1-score* de 0,9816.

A Figura 21 ilustra a pior previsão da rede neural para a segmentação de disco óptico: os resultados obtidos na imagem do Glaucoma 5. Em “A” temos a imagem pré-processada que foi enviada a rede neural para previsões; Em “B”, é a máscara valor de referência que ela deveria ter encontrado; em “C” o valor da máscara resultante da rede neural para “A”; e em “D”, a imagem enviada a rede neural com as bordas da segmentação demarcadas, sendo em verde os valores de referência e em azul a encontrada pela rede.

Figura 21 – Resultados e comparação da segmentação do disco óptico na imagem Glaucoma 5

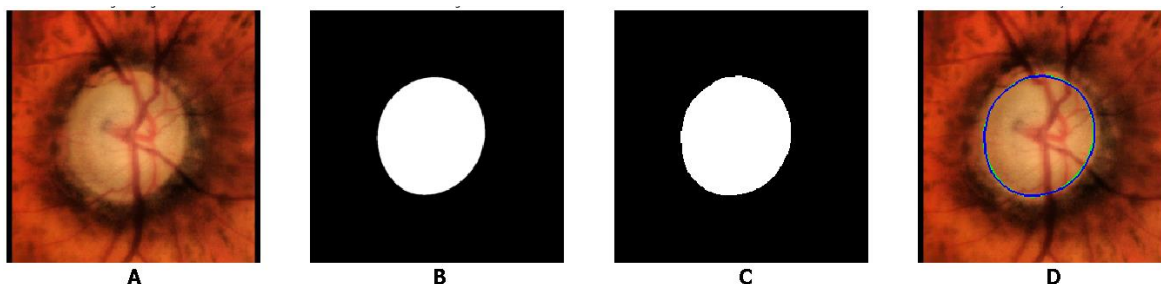


Fonte: Adaptado pelo autor de Fumero et al. (2011).

Diferente das demais imagens do *dataset*, a imagem do Glaucoma 5, que obteve o pior resultado de *F1-score*, com valor de 0,6818, possui um disco óptico com formato mais achatado e levemente inclinado em comparação com os demais, que são mais arredondados. Além disso, possui uma densa quantidade de vasos sanguíneos em relação ao tamanho total do disco óptico, principalmente na faixa central, formando assim uma grande quantidade de ruído na imagem.

O melhor resultado foi obtido na imagem Glaucoma 2, com 0,9816 de *F1-score*, ilustrada na Figura 22 onde em “A” temos a imagem pré-processada que foi enviada a rede neural para previsões; Em “B”, a máscara valor de referência que ela deveria ter encontrado; em “C” o valor da máscara resultante da rede neural para “A”; e em “D”, a imagem enviada a rede neural com as bordas da segmentação demarcadas, sendo em verde os valores de referência e em azul a encontrada pela rede.

Figura 22 – Resultados e comparação da segmentação do disco óptico na imagem Glaucoma 2



Fonte: Adaptado pelo autor de Fumero et al. (2011).

É possível identificar em “D” na Figura 22 que o resultado obtido pela rede neural é idêntico ao fornecido pelo valor de referência. Tão parecido foi o resultado que as linhas de segmentação estão praticamente sobrepostas. Este resultado foi possível devido a fatores como contraste maior entre as cores dos disco óptico, os demais componentes do olho e ao poder da rede neural U-Net em aprender os padrões de cores da imagem. Os demais resultados nas diferentes imagens ficaram bem próximos a este.

No Quadro 03 podemos ver que a média de *F1-score* obtida pela rede neural no conjunto de dados de validação foi de 0,9440, o que mostra que a rede teve capacidade de aprender muito bem a realizar a localização dos pixels do disco óptico. Visualizando por completo o resultado da segmentação de todas as 18 imagens contido no Anexo A fica ainda mais evidente o resultado positivo da rede neural para a tarefa.

O Anexo A ilustra todos os resultados obtidos pela rede neural para o procedimento de segmentação de disco óptico nas 18 imagens de fundo do olho do conjunto de teste. As bordas da segmentação foram demarcadas onde a linha verde representa a borda da segmentação dos valores de referência e, a linha azul, representa a borda da segmentação realizada pela rede neural neste estudo.

A Tabela 01 representa a média da métrica *F1-score* obtida para segmentação do disco óptico no presente trabalho em comparação com as pesquisas de Sevastopolsky (2017) e Yu et al. (2019) na base de dados RIM-ONE

R3. Na coluna da esquerda temos o nome dos autores e na direita a métrica de *F1-score* obtida.

Tabela 01 – Comparação de resultados da segmentação do disco óptico entre autores

Autor	<i>F1-score</i> médio
Sevastopolsky (2017)	0,9500
Yu et al. (2019)	0,9610
Do autor (2020)	0,9440

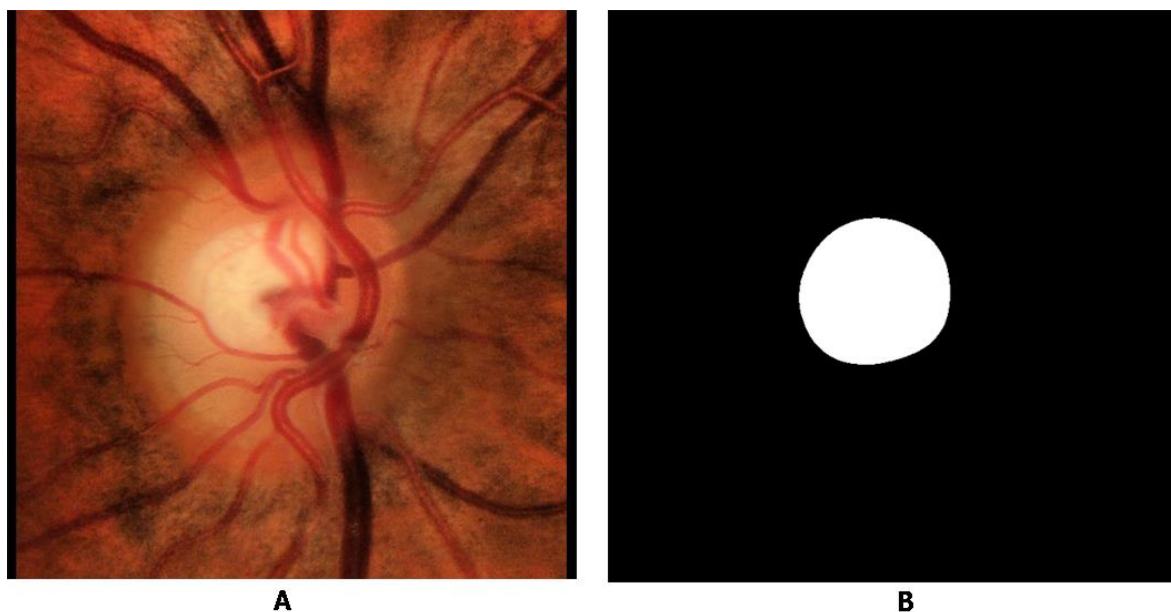
Fonte: Elaborado pelo autor (2020).

É possível observar na Tabela 01 que o resultado de *F1-score* médio obtido pelo autor, 0,9440, ficou muito próximo dos valores 0,9500 e 0,9610 obtidos por Sevastopolsky (2017) e Yu et al. (2019), respectivamente. O resultado minimamente inferior obtido pelo autor demonstra que os métodos utilizados para a segmentação do disco óptico obtidos no presente trabalho estão compatíveis com técnicas já existentes.

5.2 Segmentação da escavação

Nesta etapa é onde a rede neural aprende a realizar a segmentação da escavação. Os exemplos que são utilizados neste treinamento são os resultantes do procedimento de corte pela área de interesse. A Figura 23 representa um exemplo utilizado para treinamento da rede neural, em (a) uma imagem do fundo do olho e, em (b) sua respectiva máscara binária. Em branco são os pixels que fazem parte da escavação e em preto, os que não fazem parte.

Figura 23 – Imagem do fundo do olho cortada na área de interesse e máscara da escavação



Fonte: Adaptado pelo autor de Fumero et al. (2011).

A quantidade de etapas por época utilizada neste treinamento foi de 1.000 e a de épocas, 70. O *batch-size* foi de 2. Foram utilizadas as *callbacks ModelCheckpoint* para armazenar os resultados e *EarlyStopping* configurada para interromper o treinamento caso não ocorresse melhora dos resultados após 15 épocas.

A segmentação da escavação é uma tarefa mais desafiadora para a rede neural do que a segmentação do disco óptico, tanto para pessoas como para a rede neural. Há um contraste de cores menor comparado ao disco óptico, tornando assim uma tarefa mais difícil para a rede neural aprender.

O Quadro 04 apresenta as métricas obtidas no conjunto de dados de testes desta etapa de segmentação da escavação. Cada linha representa uma imagem e a métrica obtida. Na coluna da esquerda fica o nome da imagem informando o diagnóstico do paciente presente no *dataset* e na coluna da direita, o seu valor de *F1-score*. Quanto mais próximo de 1,00 o *F1-score* estiver, melhor o desempenho da rede. Na última linha é apresentada a média aritmética dos valores encontrados.

Quadro 04 – Resultados da segmentação da escavação na base de dados RIM-ONE R3

Nome da imagem	<i>F1-score</i>
Glaucoma 1	0,7835
Glaucoma 2	0,8596
Glaucoma 3	0,7229
Glaucoma 4	0,9654
Glaucoma 5	0,6399
Glaucoma 6	0,9280
Normal 1	0,7874
Normal 2	0,8048
Normal 3	0,8559
Normal 4	0,8133
Normal 5	0,7768
Normal 6	0,8422
Suspeito 1	0,8433
Suspeito 2	0,8969
Suspeito 3	0,8121
Suspeito 4	0,8696
Suspeito 5	0,9075
Suspeito 6	0,9200
Média	0,8350

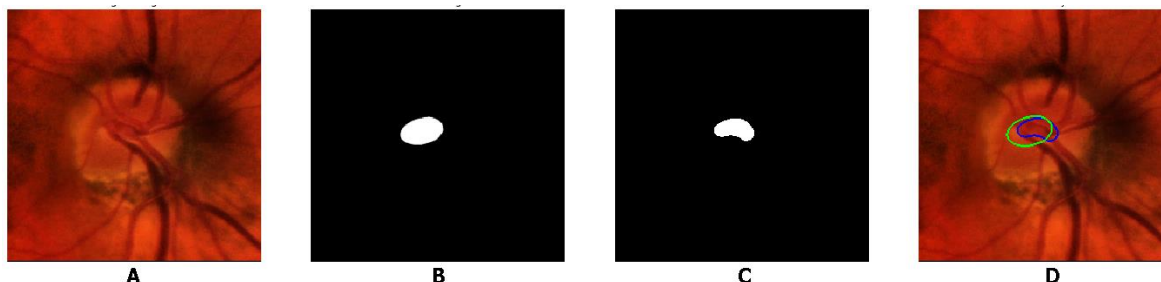
Fonte: Elaborado pelo autor (2020).

Como mostra o Quadro 04, os resultados de *F1-score* ficaram entre 0,6399 e 0,9654. Assim como na segmentação do disco óptico, a imagem Glaucoma 5 obteve o pior resultado de *F1-score*, com valor de 0,6399 para a segmentação da escavação. A imagem Glaucoma 3, com *F1-score* de 0,7229 foi o segundo pior resultado encontrado e desta vez não ficou tão distante do pior. Já o melhor resultado foi da imagem Glaucoma 4, com *F1-score* de 0,9654.

A Figura 24 ilustra a pior previsão da rede neural para a segmentação da escavação: os resultados obtidos na imagem do Glaucoma 5. Em “A” temos a imagem pré-processada que foi enviada a rede neural para previsões; Em “B”, é a

máscara valor de referência que ela deveria ter encontrado; em “C” o valor da máscara resultante da rede neural para “A”; e em “D”, a imagem enviada a rede neural com as bordas da segmentação demarcadas, sendo em verde os valores de referência e em azul a encontrada pela rede.

Figura 24 – Resultados e comparação da segmentação da escavação na imagem Glaucoma 5



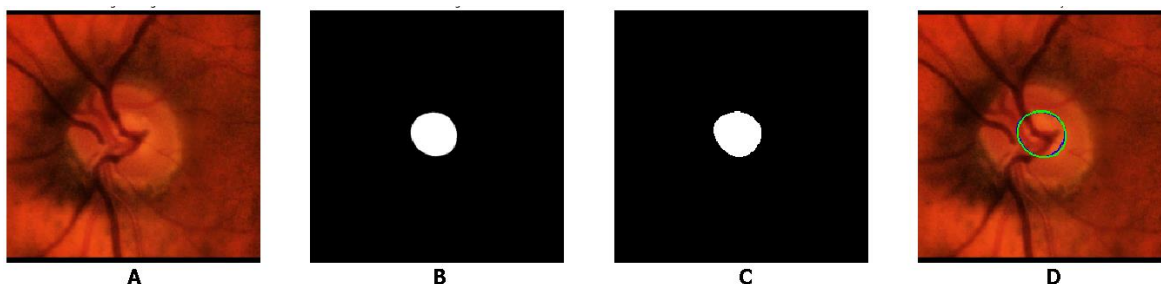
Fonte: Adaptado pelo autor de Fumero et al. (2011).

Como verificado nos resultados da segmentação do disco óptico, a imagem do Glaucoma 5, que obteve o pior resultado de *F1-score*, com valor de 0,6399, possui um disco óptico e escavação com formato distinto dos demais do *dataset*, um formato mais achatado e levemente inclinado, além da existência do ruído causado pela existência dos vasos sanguíneos.

O problema da existência dos vasos sanguíneos fica ainda mais evidente nesta etapa de segmentação. Devido a área da escavação ser pequena em relação aos vasos, e a rede neural tem dificuldade de localizar os pixels devido ao ruído por eles causado na imagem.

O melhor resultado foi obtido na imagem Glaucoma 4, com 0,9654 de *F1-score*, ilustrada na Figura 25 onde em “A” temos a imagem pré-processada que foi enviada a rede neural para previsões; Em “B”, a máscara valor de referência que ela deveria ter encontrado; em “C” o valor da máscara resultante da rede neural para “A”; e em “D”, a imagem enviada a rede neural com as bordas da segmentação demarcadas, sendo em verde os valores de referência e em azul a encontrada pela rede.

Figura 25 – Resultados e comparação da segmentação da escavação na imagem Glaucoma 4



Fonte: Adaptado pelo autor de Fumero et al. (2011).

Como ilustrado em “D” na Figura 25, é possível identificar uma comparação dos valores de referência com os valores resultantes da rede neural e concluir que ficaram bem próximos, deixando assim as linhas da borda da segmentação praticamente sobrepostas, apenas o valor encontrado, em azul, ficando um pouco mais para a esquerda do que o valor de referência.

Nota-se no Quadro 04 que a média aritmética de *F1-score* obtida pela rede neural no conjunto de dados de validação foi de 0,8350, um resultado considerado bom. Com cinco casos ficando com *F1-score* abaixo de 0,8 e quatro casos ficando com *F1-score* acima de 0,9 o resultado obtido demonstra que a rede neural foi capaz de prever com bons resultados, porém, existem alguns casos que ainda precisa de uma melhor otimização, como por exemplo, o caso demonstrado na Figura 24.

O Anexo B ilustra todos os resultados obtidos pela rede neural para o procedimento de segmentação da escavação nas dezoito imagens de fundo do olho do conjunto de teste. As bordas da segmentação foram demarcadas onde a linha verde representa a borda da segmentação dos valores de referência e, a linha azul, representa a borda da segmentação realizada pela rede neural neste estudo.

A Tabela 02 representa a média da métrica *F1-score* obtida para segmentação da escavação no presente trabalho em comparação com as pesquisas de Sevastopolsky (2017) e Yu et al. (2019) na base de dados RIM-ONE R3. Na coluna da esquerda temos o nome dos autores e na direita a métrica de *F1-score* obtida.

Tabela 02 – Comparação de resultados da segmentação da escavação entre autores

Autor	<i>F1-score</i> médio
Sevastopolsky (2017)	0,8200
Yu et al. (2019)	0,8445
Do autor (2020)	0,8350

Fonte: Elaborado pelo autor (2020).

É possível observar na Tabela 02 que o resultado de *F1-score* médio obtido pelo autor, 0,8350, ficou entre os valores 0,8200 e 0,8445 obtidos por Sevastopolsky (2017) e Yu et al. (2019), respectivamente. O método utilizado pelo autor superou aqueles utilizados por Sevastopolsky (2017) quanto a métrica *F1-score*, ficando um pouco atrás do trabalho de Yu et al. (2019), comprovando novamente a competitividade do método proposto.

6 CONCLUSÕES

O presente trabalho teve como objetivo principal verificar se era possível utilizar redes neurais convolucionais de arquitetura U-Net para realizar segmentação do disco óptico e da escavação. Para tal fim, foi desenvolvido um procedimento de preparação de imagens de fundo de olho e de implementação da arquitetura de rede neural U-Net, com o uso das bibliotecas Keras e TensorFlow, a fim de adicionar conhecimento a rede neural para que esta pudesse aprender a realizar a segmentação.

Baseando-se nos resultados obtidos, pode-se afirmar que foi possível realizar segmentação do disco óptico e da escavação com o uso de neurais convolucionais de arquitetura U-Net. Além disso, quanto à qualidade dos resultados, comparando com estudos recentes, que também tiveram resultados positivos, pode-se afirmar que o trabalho atual atinge o mesmo nível.

O grande balizador para o resultado positivo da rede neural foram os procedimentos de pré-processamento. As imagens de fundo do olho vindas diretamente do *dataset* apresentavam baixo nível de contraste entre os elementos e possuíam um grande número de dados que não eram aproveitados pela rede neural. Com o tratamento devido a estes dados a rede neural recebeu as informações de maneira simplificada, facilitando assim a obtenção de resultados positivos.

Como apresentado em resultados e discussões, na grande maioria dos casos foi possível realizar a segmentação muito próxima ao que os profissionais

realizaram, porém, quando houve presença de alguma deformação na imagem, a qualidade foi menor. A necessidade de um número maior de exemplos ficou evidente nestes casos.

Outra constatação evidenciada nos resultados foi o ruído causado pela presença vasos sanguíneos. Quanto maior a quantidade de vasos sanguíneos presentes em relação à área total, piores foram os resultados.

6.1 Sugestões de desenvolvimentos futuros

O estudo realizado mostrou que existe possibilidade de melhoria e ampliação dos resultados obtidos. São listadas a seguir algumas possibilidades de futuros avanços para uma continuidade deste trabalho.

- Ampliação da base de treinamento: Estão disponíveis na internet outros *datasets* de imagens de fundo do olho de acesso público. Para a utilização destes será necessário realizar adaptações do procedimento de pré-processamento proposto neste trabalho para que permita a utilização de diferentes imagens e máscaras de valor de referência. Esta ampliação permitirá que a rede neural aprenda características presentes em outras imagens de fundo de olho podendo assim gerar uma melhora nos resultados;
- Remoção dos vasos sanguíneos: a remoção do ruído nas imagens causado pelos vasos sanguíneos é uma técnica que pode melhorar a qualidade dos resultados. Imagens onde há menor presença de vasos tiveram melhores resultados;
- Criar interface para usuário: disponibilizar uma interface para que usuários, como por exemplo profissionais de saúde, possam utilizar a ferramenta para auxiliar na segmentação que é por eles manualmente realizada.

REFERÊNCIAS

ALMAZROA, Ahmed; BURMAN, Ritambhar; RAAHEMIFAR, Kaamran; LAKSHMINARAYANAN, Vasudevan. *Journal of Ophthalmology*, v. 2015. ***Optic disc and optic cup segmentation methodologies for glaucoma image detection: a survey***. Disponível em: <<https://www.hindawi.com/journals/joph/2015/180972/>>. Acesso em: 05 ago. 2019.

ALMAZROA et al. ***Retinal fundus images for glaucoma analysis: the RIGA dataset***. 2018. Disponível em: <<https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10579/2293584/Retinal-fundus-images-for-glaucoma-analysis-the-RIGA-dataset/10.1117/12.2293584.short?SSO=1>>. Acesso em: 01 ago. 2019.

ARMALY, Mansour. F. ***Genetic determination of cup/disc ratio of the optic nerve***. 1967. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/6067496>>. Acesso em: 01 ago. 2019.

BRAGA, Antônio de Pádua; CARVALHO, André Carlos Ponce de Leon Ferreira de; LUDERMIR, Teresa Bernarda. **Redes Neurais Artificiais: Teoria e Aplicações**. 2. ed. Rio de Janeiro: LTC, 2011.

CERENTINI, Allan. **Identificação do glaucoma em imagens do fundo do olho utilizando aprendizagem profunda**. Dissertação (Mestrado), Universidade Federal de Santa Maria. Santa Maria. 2018. Disponível em: <<https://repositorio.ufsm.br/handle/1/14008>>. Acesso em: 21 out. 2019.

CHEN, Xiangyu.; XU, Yanwu.; WONG, Damon Wing Kee; WONG, T. Y.; LIU, Jiang. ***Glaucoma detection based on deep convolutional neural network***. 2015. Disponível em: <https://www.researchgate.net/publication/308805126_Glaucoma_detection_based_on_deep_convolutional_neural_network>. Acesso em: 01 ago. 2019.

CHOLLET et al. **Keras**. 2015. Disponível em: <<https://keras.io>>. Acesso em: 21 out. 2019.

CLARK, Alex. **Pillow (PIL Fork) Documentation**. 2020. Disponível em: <<https://readthedocs.org/projects/pillow/downloads/pdf/latest/>>. Acesso em: 20 jun. 2020.

COPPIN, Ben. **Inteligência Artificial**. Rio de Janeiro: LTC, 2013.

COURROL, Lilia Coronato; PRETO, André Oliveira. **Óptica geométrica**. 1. ed. São Paulo, 2011.

FUMERO, Francisco; ALAYÓN, Silvia; SÁNCHEZ, José L.; SIGUT, José F; HERNÁNDEZ, Manuel González. **RIM-ONE: an open retinal image database for optic nerve evaluation**. 2011. Disponível em: <<https://ieeexplore.ieee.org/document/5999143>>. Acesso em: 01 out. 2019.

GARWAY-HEATH, David F.; RUBEN, Simon T.; VISWANATHAN, Ananth; HITCHINGS, Roger A. **Vertical cup/disc ratio in relation to optic disc size: its value in the assessment of the glaucoma suspect**. 1998. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/9924296>>. Acesso em: 01 ago. 2019.

GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn and TensorFlow**. 1.ed. Sebastopol. O'Reilly, 2017.

GIANGIACOMO, Annette; COLEMAN, Anne Louise. **The epidemiology of glaucoma**. 2009. Disponível em: <https://www.researchgate.net/publication/226470295_The_Epidemiology_of_Glaucoma>. Acesso em: 03 ago. 2019.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. 1.ed. MIT Press, 2016.

GOOGLE AI. **About Google AI Team**. Disponível em: <<https://ai.google/about/>>. Acesso em: 21 out. 2019.

GRAUPE, Daniel. **Principles of Artificial Neural Networks**. 3. ed. Singapore: World Scientific Publishing, 2013.

HAYKIN, Simon. **Redes Neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2008.

HAN, Jiawei; KAMBER, Micheline; PEI, Jian. **Data Mining Concepts and Techniques**. 3. ed. Elsevier, 2012.

KAPLAN, Jerry. **Artificial Intelligence: what everyone needs to know**. 1. ed. Oxford, 2016.

KARN, Ujjwal. **An Intuitive Explanation of Convolutional Neural Networks**. 2016.

Disponível em: <<https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>>. Acesso em: 21 out. 2019.

KARPATHY, Andrej. **Convolutional Neural Networks for Visual Recognition**. 2017. Disponível em: <<http://cs231n.github.io/convolutional-networks/>>. Acesso em: 21 out. 2019.

KUMAR, Sajeesh, et. al. **Glaucoma screening: analysis of conventional and telemedicine-friendly devices**. 2007. Disponível em: <https://www.researchgate.net/publication/6397007_Glaucoma_screening_Analysis_of_conventional_and_telemedicine-friendly_devices>. Acesso em: 10 set. 2019.

KUMMET, Colleen. **Statistical modeling to improve the detection of glaucoma progression**. 2013. Disponível em: <<https://www.semanticscholar.org/paper/Statistical-modeling-to-improve-the-detection-of-Kummet/ad194ddd75419e5cc12cab441a2d050b61faa7d4>>. Acesso em: 10 set. 2019.

LIMA, Alan Carlos de Moura. **Aprendizagem Profunda Aplicada ao Diagnóstico do Glaucoma**. Dissertação (Mestrado), Universidade Federal do Maranhão. São Luís, 2019. Disponível em: <<http://tede.bc.ufma.br:8080/jspui/handle/tede/2576>>. Acesso em: 1 set. 2019.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de Metodologia Científica**. 3. ed. São Paulo: Atlas, 2000.

MARY, Caroline Viola Stella; RAJSINGH, Elijah Blessing; NAIK, Ganesh. R. **Retinal fundus image analysis for diagnosis of glaucoma: A comprehensive survey**. 2016. Disponível em: <<https://ieeexplore.ieee.org/document/7536180/>>. Acesso em: 02 ago. 2019.

MEHROTRA, Kishan; MOHAN, Chilukuri K.; RANKA, Sanjay. **Elements of Artificial Neural Networks**. Massachusetts: MIT Press, 1997.

MICHELSON, Georg; WÄRNTGES, Simone; HORNEGGER, Joachim; LAUSEN, Berthold. **The papilla as screening parameter for early diagnosis of glaucoma**. 2008. Disponível em: <https://www.researchgate.net/publication/26243041_The_Papilla_as_Screening_Parameter_for_Early_Diagnosis_of_Glaucoma>. Acesso em: 21 out. 2019.

MITCHELL, Paul; SMITH, Wayne; ATTEBO, Karin; HEALEY, Paul R. **Prevalence of open angle glaucoma in Austrália: the Blue mountains eye study**. Sydney, 1996. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/8874440>>. Acesso em: 20 out. 2019.

MUELLER, John Paul; MASSARON, Luca. **Machine Learning For Dummies**. Hoboken, John Wiley & Sons, 2016.

NÄTTINEN, Janika. **Effects of Preserved and Preservative-Free Glaucoma Drugs on Proteomic Expression Levels in Corneal and Conjunctival Epithelial Cells in vitro**. Dissertação (Mestrado). Tampere, 2015. Disponível em: <<https://www.semanticscholar.org/paper/Effects-of-Preserved-and-Preservative-Free-Glaucoma-N%C3%A4ttinen-Jylh%C3%A4/38cea84fc374127683dab25fa465500fa5f07320>>. Acesso em: 02 ago. 2019.

NAYAK, Jagadish; ACHARYA, U Rajendra; BHAT, P Subbanna; SHETTY, Nakul; LIM, Teik-Cheng. **Automated diagnosis of glaucoma using digital fundus images**. 2009. Disponível em: <https://www.researchgate.net/publication/38009750_Automated_Diagnosis_of_Glaucoma_Using_Digital_Fundus_Images>. Acesso em: 01 ago. 2019.

OPENCV, *Open Source Computer Vision Library*. **Opencv Documentation**. 2020. Disponível em: <<https://opencv.org/about/>>. Acesso em: 20 jun. 2020.

PASCOLINI, Donatella; MARIOTTI, Silvio Paolo. **Global estimates of visual impairment: 2010**. Genebra, 2011. Disponível em: <<https://www.who.int/blindness/publications/globaldata/en/>>. Acesso em: 23 out. 2019.

PASSOS, Erinaldo Costa.; ANDRADE-NETO, Antonio Vieira.; LEMAIRE, Thierry. **Comportamento Ótico do olho humano e suas ametropias**. Feira de Santana, 2008. Disponível em: <<http://dfisweb.uefs.br/caderno/vol6n12/Olhohumano.pdf>>. Acesso em: 23 out. 2019.

PEDREGOSA, F. et al. **Scikit-learn: Machine Learning in Python**. *Journal of Machine Learning Research*. p. 2825-2830. 2011.

PYTHON; **General Python FAQ –Python 3.6 documentation**. Disponível em: <<https://docs.python.org/3/faq/general.html>>. Acesso em: 21 out. 2019.

RAGHAVENDRA U. et al. **Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images**. 2018. Disponível em: <https://www.researchgate.net/publication/322892996_Deep_Convolution_Neural_Network_for_Accurate_Diagnosis_of_Glaucoma_Using_Digital_Fundus_Images>. Acesso em: 02 ago. 2019.

RAMOS, Albenides. **Metodologia da pesquisa científica: como uma monografia pode abrir o horizonte do conhecimento**. São Paulo: Atlas, 2009.

RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. **U-Net: Convolutional Networks for Biomedical Image Segmentation**. Alemanha,

2015. Disponível em: <<https://arxiv.org/abs/1505.04597>>. Acesso em: 20 out. 2019.

RUENGKITPINYO, Wanicha; VEJJANUGRAHA, Pikul; KONGPRAWECHNON, Waree; KONDO, Toshiaki; BUNNUN, Pished; KANEKO, Hirohiko. ***An automatic glaucoma screening algorithm using cup-to-disc ratio and ISNT rule with support vector machine***. Yokohama, 2015. Disponível em: <<https://ieeexplore.ieee.org/document/7392152>>. Acesso em: 01 set. 2019.

SANTOS, Antonio Raimundo dos. **Metodologia científica: a construção do conhecimento**. 2. ed. Rio de Janeiro: DP&A editora, 1999.

SETIAWAN, Agung W.; MENGKO, Tati R.; SANTOSO, Oerip S.; SUKSMONO, Andriyan B. ***Color retinal image enhancement using CLAHE***. Jakarta: Indonésia. 2013. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/6588092>>. Acesso em: 06 jun. 2020.

SEVASTOPOLSKY, Artem. ***Optic Disc and Cup Segmentation Methods for Glaucoma Detection with Modification of U-Net Convolutional Neural Network***. Moscou, 2017. Disponível em: <<https://link.springer.com/article/10.1134/S1054661817030269>>. Acesso em: 02 ago. 2019.

SHALLUE, Christopher J.; VANDERBURG, Andrew. ***Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90***. Chicago, 2018. Disponível em: <<https://iopscience.iop.org/article/10.3847/1538-3881/aa9e09>>. Acesso em:

SIVASWAMY et al. ***A Comprehensive Retinal Image Dataset for the Assessment of Glaucoma from the Optic Nerve Head Analysis***. Índia, 2014. Disponível em: <<https://ieeexplore.ieee.org/document/6867807>>. Acesso em: 01 ago. 2019.

TAN, Pang – Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao Datamining: mineração de dados**. 1. ed. Rio de Janeiro: Editora Ciência Moderna, 2009.

VAN DER WALT, Stéfan; SCHÖNBERGER, Johannes L.; NUNEZ-IGLESIAS, Juan; BOULOGNE, François; WARNER, Joshua D.; YAGER, Neil; GOUILLART, Emmanuelle; YU, Tony. ***Scikit-image: image processing in Python***. 2014. Disponível em: <<https://peerj.com/articles/453/>>. Acesso em: 22 jun. 2020.

VAN DER WALT, Stéfan; COLBERT, S. Chris; VAROQUAUX, Gael. ***The NumPy Array: A Structure for Efficient Numerical Computation. Computing in Science & Engineering***. v. 13. 2011. Disponível em: <<https://ieeexplore.ieee.org/document/5725236/authors>>. Acesso em: 20 jun. 2020.

VENNERS, Bill. ***The Making of Python - A Conversation with Guido van Rossum, Part I***. 2003. Disponível em: <<https://www.artima.com/intv/python.html>>. Acesso em 21 out. 2019.

WALTERS, Austin Grant. ***Convolutional Neural Networks (CNN) to Classify Sentences***. 2019. Disponível em: <<https://austingwalters.com/convolutional-neural-networks-cnn-to-classify-sentences/>>. Acesso em: 23 out. 2019.

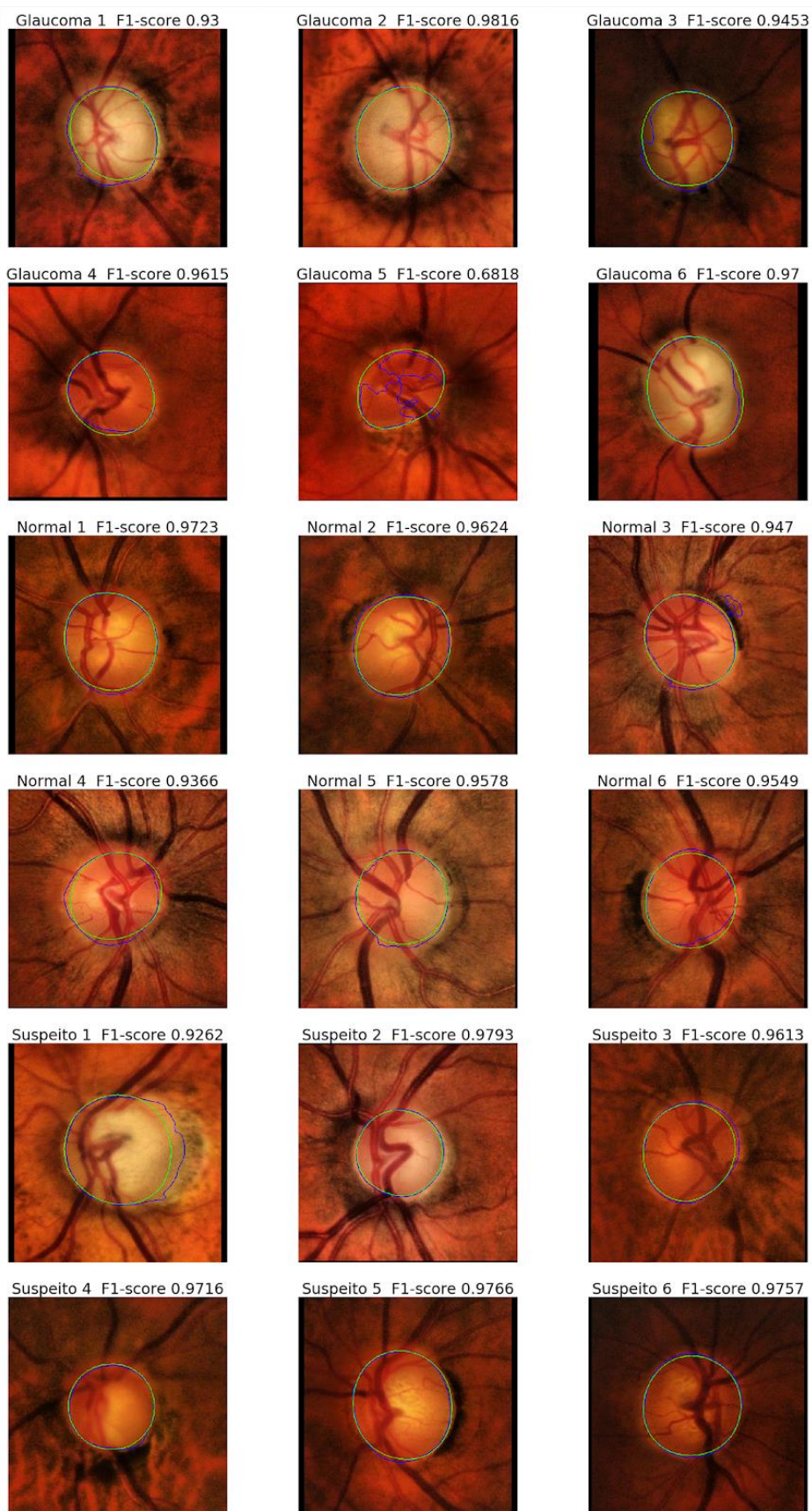
WEINREB, Robert N.; KHAW, Peng Tee. ***Primary open-angle glaucoma***. 2004. Disponível em: <[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(04\)16257-0](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(04)16257-0)>. Acesso em: 26 out. 2019.

YANG, Chunlan; LU, Min; DUAN, Yanhua; LIU, Bing. ***An efficient optic cup segmentation method decreasing the influences of blood vessels***. 2018. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6158914/>>. Acesso em: 30 mai. 2020.

YU, Shuang; XIAO, Di; FROST, Shaun; KANAGASINGAM, Yogesan. ***Robust optic disc and cup segmentation with deep learning for glaucoma detection***. 2019. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0895611118305573>>. Acesso em: 01 ago. 2019.

ANEXOS

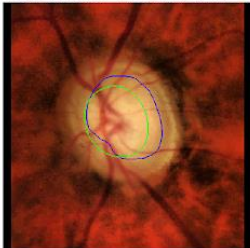
ANEXO A – Resultados da segmentação do disco óptico



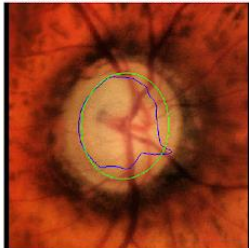
Fonte: Adaptado pelo autor de Fumero et al. (2011).

ANEXO B – Resultados da segmentação da escavação

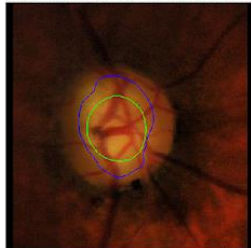
Glaucoma 1 F1-score 0.7747



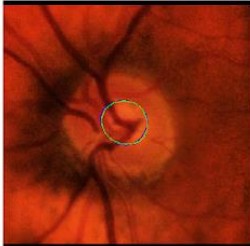
Glaucoma 2 F1-score 0.8673



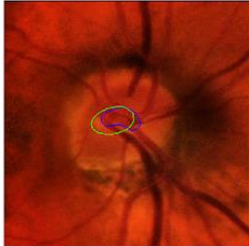
Glaucoma 3 F1-score 0.7029



Glaucoma 4 F1-score 0.97



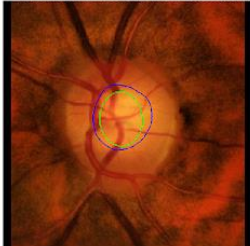
Glaucoma 5 F1-score 0.6518



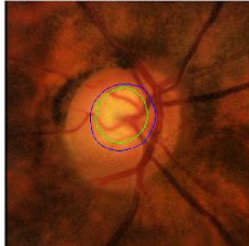
Glaucoma 6 F1-score 0.927



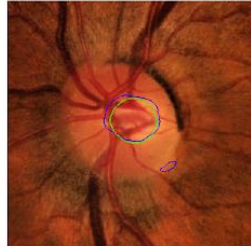
Normal 1 F1-score 0.7576



Normal 2 F1-score 0.7805



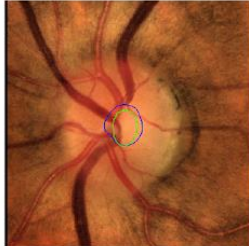
Normal 3 F1-score 0.8418



Normal 4 F1-score 0.8261



Normal 5 F1-score 0.7405



Normal 6 F1-score 0.8202



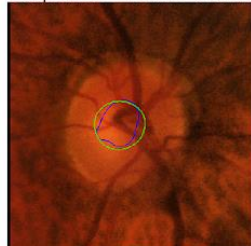
Suspeito 1 F1-score 0.8401



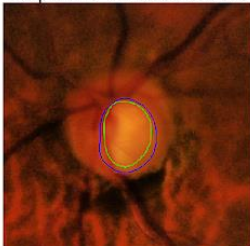
Suspeito 2 F1-score 0.8833



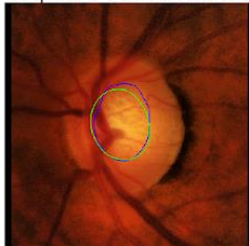
Suspeito 3 F1-score 0.8319



Suspeito 4 F1-score 0.8462



Suspeito 5 F1-score 0.9095



Suspeito 6 F1-score 0.9098



Fonte: Adaptado pelo autor de Fumero et al. (2011).



UNIVATES

R. Avelino Talini, 171 | Bairro Universitário | Lajeado | RS | Brasil
CEP 95914.014 | Cx. Postal 155 | Fone: (51) 3714.7000
www.univates.br | 0800 7 07 08 09